UTILITY

A P P L I C A T I O N

for

UNITED STATES LETTERS PATENT

on

METHODS FOR IDENTIFYING CANCER RISK

by

Andrew P. Feinberg

Docket No.: **JHU1990-1**
Drawings: **Nine Sheets**

Lisa Haile, JD, Ph.D.
GRAY CARY WARE & FREIDENRICH LLP
4365 Executive Drive, Suite 1100
San Diego, California 92121-2133

METHODS FOR IDENTIFYING CANCER RISK

RELATED APPLICATION DATA

[0001] This application claims the benefit of priority under 35 U.S.C. § 119(e) of U.S. Serial No. 60/398,660, filed July 26, 2002, the entire contents of which is incorporated herein by reference. This application is a continuation-in-part application of U.S. Serial No. 10/336,552, filed on January 3, 2003.

STATEMENT OF GOVERNMENT SUPPORT

[0002] This invention was made in part with government support under Grant No. R01 CA65145 and K07 CA092445 awarded by the National Institutes of Health. The United States government has certain rights in this invention.

BACKGROUND OF THE INVENTION

FIELD OF THE INVENTION

[0003] The present invention relates generally to methods for detecting the presence of or risk of developing cancer and more specifically to methods for detecting the presence of hypomethylation of the H19 gene and the IGF2 gene.

BACKGROUND INFORMATION

[0004] The single greatest impediment to cancer diagnosis is the general requirement that the tumor itself must be detected directly. Efforts to identify genetic abnormalities in normal tissues of patients with cancer or at risk of cancer have been disappointing. For example, BRCA1 mutations are present in only about 1% of breast cancers. A small fraction of patients with colorectal cancer have predisposing mutations in the APC gene (>1%), causing adenomatous polyposis coli. An even smaller fraction show mutations in genes responsible for replication error repair (>2% of colon cancer patients, or much less than 1%

of the population), show mutations in genes responsible for nucleotide mismatch error repair causing hereditary nonpolyposis colorectal cancer (HNPCC or Lynch syndrome).

[0005]   Diagnostic methods such as microsatellite instability, require for identification that a patient already have a tumor. For example, microsatellite instability compares microsatellite marker length between the monoclonal tumor cell population and normal tissue derived from the same patient.

[0006]   Family history still remains the most reliable diagnostic procedure for identifying patients at risk of cancer. A molecular diagnostic approach that might identify patients with cancer or at risk of cancer, using only normal tissue, would offer a decisive advantage for intervention and treatment.

[0007]   Except for rare hereditary cancer syndromes, the impact of molecular genetics on cancer risk assessment and prevention has been minimal. Cancer surveillance has been effective for some cancers in which risk can be identified, for example colorectal cancer in familial adenomatous polyposis coli and hereditary nonpolyposis colorectal cancer (Markey, L., et al., *Curr. Gastroenterol. Rep.* **4**, 404-413 (2002)), but these syndromes cumulatively account for less than 1% of cancer patients (Samowitz, W. S., et al., *Gastroenterology* **121**, 830-838 (2001); Percesepe, A., et al., *J. Clin. Oncol.* **19**, 3944-3950 (2001)). Nevertheless, genetics is thought to contribute substantially to cancer risk, since the odds ratio for malignancy increases in patients with first degree relatives with cancer, e.g., 2 to 3-fold in colorectal cancer (Fuchs, C. S., et al., *N. Engl. J Med.* **331**, 1669-1674 (1994)). Therefore, there remains a need to develop genetic tests to identify these patients.

[0008]   Accordingly, no tests are available for identifying common cancer risk in the general population. As discussed above, genetic abnormalities that are known to predispose to cancer are rare. At the same time, advances in cancer treatment have had a small impact on morbidity and mortality. A major advance in cancer requires identification of patients at risk (i.e. identifies patients before they develop cancer), which could be combined with increased surveillance and chemoprevention, similar to the modern approach to cardiovascular medicine.

[0009] Thus, there remains a need for a diagnostic method for detecting and/or screening for the presence of diseases and/or the risk of developing a disease. In particular, there remains a need for a method for detecting and/or screening for the presence of cancer, for example colorectal cancer. There also remains a need for a method of detecting and/or screening for the presence of cancer and/or the risk of developing cancer that can be applied to a wide section of the population.

[0010] Epigenetic alterations in human cancer, i.e., alterations in the genome other than the DNA sequence itself, were first described in 1983 by Feinberg and Vogelstein (A.P. Feinberg et al., *Nature* (Lond.), 301: 89-92, 1983), who found widespread hypomethylation of genes in CRCs and in premalignant adenomas. Epigenetic abnormalities identified subsequently include global genomic hypomethylation (A.P. Feinberg et al., *Cancer Res.*, 48: 1159-1161, 1988), promoter hypermethylation of CpG islands (S.B. Baylin, et al., *Cancer Res.*, 46: 2917-2922, 1986; A. Merlo et al, *Nat. Med.*, 1: 686-692, 1995), and LOI (Rainier S. Johnson et al, *Nature* (Lond.), 362: 747-749, 1993; O. Ogawa et al., *Nature* (Lond.), 362: 749-751, 1993), or loss of the normal parent of origin-dependent gene silencing, affecting at least the genes IGF2, PEG1, p73, and LIT1 (S. Rainier et al., *Nature* (Lond.), 362: 747-749, 1993; O. Ogawa et al., *Nature* (Lond.), 362: 749-751, 1993; I.S. Pedersen et al, *Cancer Res.*, 59: 5449-5451, 1999; M.P. Lee et al, *Proc. Natl. Acad. Sci.* USA, 96: 5203-5208, 1999; M. Kohda et al, *Mol. Carcinog.*, 31: 184-191, 2001; Y.C., Cai et al., *Carcinogenesis* (Lond.), 21: 683-689, 2000; K. Tanaka et al, *Oncology*, 60: 268-273, 2001). LOI of IGF2 causes overexpression of IGF2 (J.D. Ravenel et al, *J. Natl. Cancer Inst.*, 93: 1698-1703, 2001), an important autocrine growth factor in cancer. LOI was first identified in embryonal tumors in childhood, including Wilms' tumor, in which it is the most common molecular alteration (S. Rainier et al, *Nature* (Lond.), 362: 747-749, 1993; O. Ogawa et al, *Nature* (Lond.), 362: 749-751, 1993), as well as rhabdomyosarcoma (S. Zhan et al, *J. Clin. Investig.*, 94: 445-448, 1994) and hepatoblastoma (S. Rainier et al, *Cancer Res.*, 55: 1836-1838, 1995). LOI was also later found in common adult malignancies including ovarian (H.T. Kim et al, *Am. J. Med. Genet.*, 80: 391-395, 1998), colon (H. Cui et al, *Nat. Med.*, 4: 1276-280, 1998), lung (M. Kondo et al, *Oncogene*, 10: 1193-1198, 1995), and bladder cancer (M. Elkin et al, *FEBS Lett.*, 374: 57-61, 1995), as well as chronic

myelogenous leukemia (G.S. Randhawa et al, *Blood*, 91: 3144-3147, 1998). In CRC, LOI
is particularly important because it is found commonly in both the tumor and normal tissue
of patients with CRC, at ~3-fold higher frequency then in patients without colon tumors (H.
Cui et al, *Nat. Med.*, 4: 1276-280, 1998), and, thus, LOI may represent the only common
alteration linked to cancer that is found in normal tissue.

[0011]   In Wilms' tumors, approximately half of tumors appear to arise by an epigenetic
mechanism involving LOI rather than genetic alterations involving, for example, WT1
mutations and LOH, and the tumors with LOI appear in children who develop cancer at a
later age, accounting for the bimodal age distribution of Wilms' tumor (J.D. Ravenel et al,
J. Natl. Cancer Inst., 93: 1698-1703, 2001). LOI was linked to increased methylation,
because Wilms' tumors with LOI of IGF2, i.e., activation of the normally silent maternal
allele, show aberrant methylation of the normally unmethylated maternal allele of a DMR
upstream of the H19 gene on the same chromosome (M.J. Steenman et al, *Nat. Genet.*, 7:
433-439, 1994; T. Moulton et al, *Nat. Genet.*, 7: 440-447, 1994). This result is consistent
with the enhancer competition model for regulation of H19 imprinting. By this model,
IGF2 and H19 promoters compete on the same chromosome for a shared enhancer, and
access of the maternal IGF2 allele to this enhancer is blocked by the H19 DMR when
unmethylated, likely because of the insulator activity of CTCF binding to the unmethylated
H19 DMR (P.A. Leighton et al, *Nature* (Lond.), 375: 34-39, 1995; R. Ohlsson et al, *Trends
Genet.*, 17: 520-527, 2001; W. Reik et al, *Nature* (Lond.), 405: 408-409, 2000; A.T. Hark et
al, *Nature* (Lond.), 405: 486-489, 2000; A.C. Bell et al, *Nature* (Lond.), 405: 482-485,
2000). Consistent with this, it has been observed that in Wilms' tumor, methylation of the
maternal H19 DMR includes CTCF-binding sites (H. Cui et al, *Cancer Res.*, 61: 4947-4950,
2001). These results would suggest that increased or ectopic activity of a DNA
methyltransferase might lead to aberrant methylation of the maternal H19 DMR.

[0012]   Therefore, it was surprising to observe that HCT116, a CRC line with normal
imprinting of IGF2, is hypermethylated at H19 and retains normal imprinting after somatic
cell knockout of the maintenance DNA methyltransferase DNMT1 but loses imprinting
after subsequent somatic cell knockout of DNMT3B (*Nature* (Lond.), 416: 552-556, 2002),

a *de novo* methyltransferase, i.e., that is able to methylate unmethylated sequences and is necessary for normal imprinting (M. Okano et al, *Cell*, 99: 247-257, 1999; K. Hata et al, *Development* (Camb.), 129: 1983-1993, 2002). Therefore, there remains a need to determine the relationship between methylation state of IGF2 and H19 and loss of imprinting and/or cancer risk, such as colorectal cancer risk, and to devise methods for identifying cancer risk based on this relationship

[0013]   The results described herein differ from past studies, and suggest a model of IGF2 imprinting in at least the colon that differs from the conventional view of enhancer competition between IGF2 and H19.

SUMMARY OF THE INVENTION

[0014]   The present invention provides methods and kits that are based on the finding of an association between loss of imprinting (LOI) and family history of colorectal cancer (CRC) and between LOI and present or past personal history of colorectal neoplasia. Furthermore, the present invention provides methods, kits, and nucleic acid primers, that are based on the finding that hypomethylation of both the IGF2 gene and the H19 gene is correlated with loss if imprinting of the IGF2 gene. Loss of imprinting of IGF2 is correlated with the presence and increased risk for developing cancer, especially colorectal cancer.

[0015]   Accordingly, methods of the present invention analyze LOI, especially LOI of the IGF2 gene, for example by analyzing hypomethylation of the IGF2 gene or H19 gene, to identify an increased risk of developing cancer in a subject.

[0016]   Accordingly, in one embodiment, the present invention provides a method for identifying loss of imprinting of the IGF2 gene in a subject. The method in this embodiment includes analyzing a biological sample from the subject for hypomethylation of a differentially methylated region (DMR) of the H19 gene and/or the IGF2 gene, or a polymorphism and/or fragment of the H19 DMR and/or IGF2 DMR. Typically, the subject is a human subject. Furthermore, in certain aspects, the H19 DMR or fragment thereof, includes a CTCF binding site, for example, CTCF binding site 1 or CTCF binding site 6.

[0017] In certain aspects of this embodiment, the subject is an apparently normal subject. Furthermore, in certain aspects, hypomethylation is analyzed in a DNA region corresponding to SEQ ID NO:6, an H19 DMR. In certain aspects the method comprises analyzing the biological sample for hypomethylation of positions within the region of the H19 DMR that are analyzed using the nested primer pairs SEQ ID NOs:23 and 24, followed by SEQ ID NOs:25 and 26, as illustrated in the Examples herein. Furthermore, in certain aspects, hypomethylation is analyzed in a DNA region corresponding to SEQ ID NO:1, an IGF2 DMR. In certain aspects the method comprises analyzing the biological sample for hypomethylation of positions within the region of the IGF2 DMR that are analyzed using the nested primer pairs SEQ ID NOs:2 and 3, followed by SEQ ID NOs:4 and 5, or the region analyzed using primer pairs SEQ ID NOs: 27 and 28, followed by SEQ ID NOs:29 and 30, as illustrated in the Examples herein.

[0018] In another embodiment, the present invention relates to a method for identifying an increased risk of developing cancer in a subject. The method includes analyzing a biological sample from the subject for hypomethylation of a differentially methylated region (DMR) of an H19 gene, or a polymorphism and/or fragment of the H19 DMR. In certain aspects of the invention, the cancer is colorectal cancer. In certain aspects, the method further includes analyzing the biological sample for hypomethylation of a DMR of the IGF2 gene, SEQ ID NO:1, or a polymorphism and/or fragment thereof.

[0019] In another embodiment, the present invention provides a method for identifying an increased risk of developing cancer in a subject. The method includes analyzing a biological sample from the subject for loss of imprinting of the IGF2 gene. According to the method a loss of imprinting is indicative of an increased risk of developing cancer. In certain embodiments, the method includes analyzing the genomic DNA for hypomethylation of a differentially methylated region (DMR) of IGF2 and/or a DMR of H19, or a polymorphism and/or fragment of the IGF2 DMR and/or the H19 DMR. The region analyzed in one aspect of the invention corresponds to SEQ ID NO:1 or a polymorphism thereof, or SEQ ID NO:6, or a polymorphism thereof. The method can be performed during routine clinical care, on a subject having no apparent or suspected

hyperproliferative disorder such as cancer. The first biological sample can be a blood sample, for example.

[0020]   In one aspect, the method can further include analysis of a second biological sample from the subject at a target tissue for loss of imprinting of the IGF2 gene, wherein a loss of imprinting in the second sample is indicative of an increased risk of developing cancer in the target tissue. In certain embodiments, the second biological sample is not a blood sample. For example, the first biological sample can be a blood sample and the second biological sample can be isolated from colorectal tissue. In embodiments where the second biological sample is isolated from colorectal tissue, the cancer is typically colorectal cancer.

[0021]   In another embodiment, the present invention provides a method for managing health of a subject. The method includes performing the method discussed above, and then performing a traditional cancer detection method on the subject if the subject has an increased risk for developing cancer. The traditional cancer detection method can be, for example, colonoscopy, especially where the subject has an increased risk of developing colorectal cancer.

[0022]   In yet another embodiment, the present invention provides a method for prognosing cancer risk of a subject. The method includes analyzing a first biological sample from the subject for altered methylation of a differentially methylated region (DMR) of the IGF2 gene and/or of the H19 gene, or a polymorphism and/or fragment of the IGF2 DMR and/or the H19 DMR. In certain aspects, the method analyzes SEQ ID NO:1 or SEQ ID NO:6. Hypomethylation is indicative of an increased risk of developing cancer. In this embodiment, the first biological sample is typically a blood sample.

[0023]   In one embodiment, the present invention provides a method for identifying predisposition to colorectal cancer of a subject. The method includes identifying a loss of imprinting in a biological sample from the subject and correlating the loss with a predisposition to colorectal cancer. Loss of imprinting is associated with an increased predisposition to colorectal cancer.

[0024]   In another embodiment, the present invention provides a kit for determining a methylation status of a differentially methylated region (DMR) of IGF2 corresponding to SEQ ID NO:1 or a polymorphism or fragment thereof, or of the DMR of H19, corresponding to SEQ ID NO:6, or a polymorphism or fragment thereof. The kit includes an oligonucleotide probe, primer, or primer pair, or combination thereof, capable of binding to the DMR with or without prior bisulfite treatment of the DMR. In certain aspects, the kit includes one or more detectable labels. ·

[0025]   The kit, in certain aspects, also includes a plurality of oligonucleotide probes, primers, or primer pairs, or combinations thereof, capable of binding to the DMR of IGF2 or H19 with or without prior bisulfite treatment of the DMR. The kit can include an oligonucleotide primer pair that hybridizes under stringent conditions to all or a portion of the DMR only after bisulfite treatment. The kit can include instructions on using kit components to identify an increased risk of developing cancer. In certain embodiments the instructions are directed at subjects of the general population. The kit for example, includes one or both of a primer pair corresponding to the primer pair SEQ ID NO:23 and SEQ ID NO:24 and the primer pair SEQ ID NO: 25 and SEQ ID NO:26. In another aspect, the kit for example, includes one or both of a primer pair corresponding to the primer pair SEQ ID NO:27 and SEQ ID NO:28, and the primer pair SEQ ID NO: 29 and SEQ ID NO:30.

[0026]   In another embodiment, the present invention provides isolated oligonucleotides and primer pairs corresponding to SEQ ID NO:23 and SEQ ID NO:24; SEQ ID NO:25 and SEQ ID NO:26; SEQ ID NO:27 and SEQ ID NO:28; and SEQ ID NO:29 and SEQ ID NO:30, and isolated nucleic acids that correspond to a region of the IGF2 or H19 DMRs that is amplified by SEQ ID NO:23 and SEQ ID NO:24; SEQ ID NO:25 and SEQ ID NO:26; SEQ ID NO:27 and SEQ ID NO:28; and SEQ ID NO:29 and SEQ ID NO:30.

BRIEF DESCRIPTION OF THE DRAWINGS

[0027]   Figure 1 provides the sequence of a differentially methylated region (DMR) of the IGF2 gene. The sequence corresponds to SEQ ID NO:1. The IGF2 DMR corresponds

to residues −566 bp to −311 bp relative to human IGF2 exon 3, which corresponds to positions 661 to 916 of GenBank accession no. Y13633.

[0028] Figures 2A and 2B show methylation status and imprinting analysis of the IGF2 DMR of SEQ ID NO:1, in PBL and normal colonic mucosa with either normal imprinting or LOI of IGF2. (A) IGF2 DMR shows normal methylation in PBL and normal colonic mucosa with normal imprinting of IGF2. Shown are the methylation states of individual PCR products subcloned after bisulfite treatment and PCR. In some cases, individual alleles can be distinguished by single nucleotide polymorphisms on the same PCR product (shown at right). Ten to 20 clones were sequenced depending on heterozygosity at the SNP site. Filled circles represent methylated cytosine and open circles represent unmethylated cytosine. Imprinting analysis was performed by hot-stop PCR (Uejima, H., et al., *Nat. Genet.* **25**, 375-376 (2000)), and is shown without (−) and with (+) reverse transcriptase. LOI index is displayed numerically (LOI index = (less active allele/ more active allele) X 100%). LOI is defined as an LOI index > 25 (Cui, H., et al., *Nat. Med.* **4**, 1276-1280 (1998)). Patient samples are matched PBL (Patient number includes "L") and colon from the same individuals (Patent number includes "C"). Single nucleotide polymorphisms distinguishing alleles are shown on the left. (B) IGF2 DMR shows biallelic hypomethylation in PBL and normal colonic mucosa with LOI of IGF2, except for one sample that displays partial methylation of both alleles.

[0029] Figure 3 provides the sequence of a differentially methylated region (DMR) of the H19 gene. The sequence corresponds to SEQ ID NO:6 The H19 DMR corresponds to nucleotides 2057 to 8070 of Genbank accession no. AF087017 (SEQ ID NO:37).

[0030] Figure 4 shows hypomethylation of H19 and IGF2 DMRs in CRC cell lines with LOI of IGF2 and somatic cell knockout of DNA methyltransferase. HCT116 cells (WT), DNMT1 knockout (T1KO), DNMT3B knockout (3BKO), and DNMT1/DNMT3B double-knockout (DKO) cells were analyzed by bisulfite genomic sequencing at the H19 (CTCF binding site 1) and IGF2 DMRs. By the enhancer competition model, biallelic methylation of the H19 DMR should cause LOI of IGF2, but it does not in these cells. Rather, the IGF2

DMR shows normal half methylation with normal imprinting and hypomethylation with LOI of IGF2.

[0031] Figure 5 shows normal methylation of IGF2 and variable methylation of H19 DMRs in sporadic CRCs with normal imprinting of IGF2. The H19 DMR shows a general trend of normal methylation and normal imprinting status of IGF2, whereas the IGF2 DMR shows complete concordance of normal methylation and normal imprinting. Genomic DNA was treated with sodium bisulfite, and then was PCR amplified and subcloned before sequencing. Ten to 15 clones were sequenced for each sample. Each line represents a separate clone. (·), methylated CpG sites; (o), unmethylated CpG sites. Case number is shown on the left. Single nucleotide polymorphisms are shown on the left, distinguishing alleles.

[0032] Figure 6 shows hypomethylation of H19 and IGF2 DMRs in sporadic CRCs with LOI. The H19 DMR shows a general trend of hypomethylation and LOI of IGF2, whereas the IGF2 DMR shows complete concordance of hypomethylation and LOI. Labeling is as described in the legend to Figure 5.

DETAILED DESCRIPTION OF THE INVENTION

[0033] Epigenetic alterations in human cancers include global DNA hypomethylation, gene hypomethylation and promoter hypermethylation, and loss of imprinting (LOI) of the insulin-like growth factor-II gene (IGF2). A mechanism for LOI described previously is hypermethylation of a differentially methylated region (DMR) upstream of the HI9 gene, allowing activation of the normally silent maternal allele of IGF2. The present specification discloses that this mechanism does not apply to colorectal cancers, which show hypomethylation of the H19 DMR as well as the DMR upstream of exon 3 of IGF2. This hypomethylation is found in both colorectal cancers and normal mucosa from the same patients, and in cell lines with somatic cell knockout of DNA methyltransferases DNMT1 and DNMT3B. These data suggest that hypomethylation is a mechanism for LOI, that the popular IGF2-H19 enhancer competition model for IGF2 imprinting does not apply to the

human colon, and that an alternative model for LOI would involve a transcriptional repressor acting on the normally silent maternal allele of IGF2.

[0034] Accordingly, in one embodiment, the present invention provides a method for identifying loss of imprinting of the IGF2 gene in a subject. The method in this aspect includes analyzing a biological sample from the subject for hypomethylation of a differentially methylated region (DMR) of the H19 gene and/or the IGF2 gene, and/or a polymorphism and/or fragment of the H19 DMR and/or the IGF2 DMR.

[0035] Typically, the subject is a human subject. Furthermore, in certain aspects, the H19 DMR or portion thereof, includes a CTCF binding site, for example, CTCF binding site 1 or CTCF binding site 6.

[0036] In certain aspects of this embodiment, the subject is an apparently normal subject. Therefore, in certain aspects, the subject is not known to have a colorectal neoplasm. Furthermore, in certain aspect, hypomethylation is analyzed in a DNA region corresponding to SEQ ID NO:6, an H19 DMR.

[0037] In certain aspects, the method further includes analyzing the biological sample for hypomethylation of a DMR of the IGF2 gene that corresponds to SEQ ID NO:1.

[0038] In one embodiment, the present invention relates to a method for identifying an increased risk of developing cancer in a subject. The method includes analyzing a biological sample from the subject for hypomethylation of a differentially methylated region (DMR) of an H19 gene, or a polymorphism and/or fragment of the H19 DMR. In certain aspects of the invention, the cancer is colorectal cancer.

[0039] One IGF2 DMR corresponds to SEQ ID NO:1, and is located at position −566 bp to −311 bp relative to exon 3 of IGF2 (i.e., nucleotides 661 to 916 of GenBank accession no. Y13633 (SEQ ID NO:35), incorporated herein in its entirety by reference). The DMR of SEQ ID NO:1 is analyzed for hypomethylation in preferred embodiments of the present invention. Other DMRs have been reported for the IGF2 gene, for example, a DMR around exon 3 (Sullivan MJ et al., *Oncogene,* (1999) 18(52):7527-34.). Residues which are

methylated within SEQ ID NO:1, typically CpG residues, can be identified using methods
such as bisulfite genomic sequencing, as known in the art and as discussed in further detail
herein (See e.g., Vu et al., *Genomics*, 64:132-43 (2000), incorporated herein in its entirety
by reference). In preferred embodiments, methods of the present invention analyze 3 CpG
residues displaying differential methylation that correspond to positions 747, 750 and 766
of the Genbank entry Y13633, incorporated herein in its entirety by reference (SEQ ID
NO:35) (See FIGS. 4-6).

[0040]   One DMR of H19 corresponds to SEQ ID NO:6 (i.e., nucleotides 2057 to 8070 of
Genbank accession no. AF087017, incorporated herein by reference in its entirety (SEQ ID
NO:37); which correspond in variant form to nucleotides 3829 to 9842 of AF125183 (SEQ
ID NO:36) (See Cui et al., *Canc. Res.*, 61:4947-4950 (2001), incorporated herein by
reference in its entirety; See also Cui et al., *Canc. Res.*, 62:6442-6446 (2002), incorporated
herein by reference in its entirety)). The DMR of H19 is analyzed for hypomethylation in
preferred embodiments of the present invention. Furthermore, it is known that DMRs are
found in the human in all seven CTCF binding sites that showed differential methylation
upstream of H19 gene. In addition, the H19 promoter region also has been shown to be
differentially methylation.(Hark AT et al, *Nature*. (2000);405(6785):486-9., Bell AC et al.,
Nature. (2000) 405(6785):482-5. Steenman MJ et al., *Nat Genet*. (1994) 7(3):433-9.).
Residues which are methylated within SEQ ID NO:6, typically CpG residues, can be
identified using methods such as bisulfite genomic sequencing, as known in the art and as
discussed in further detail herein (See e.g., Vu et al., Genomics, 64:132-43 (2000)). All of
the CpG residues within SEQ ID NO:6 show differential methylation (See FIGS. 4-6).

[0041]   It will be recognized that in certain aspects, methods of the invention can
determine methylation status by utilizing primers that bind the IGF2 gene or H19 gene at
regions outside of the IGF2 DMR or the H19 DMR, but that allow analysis of a region of
the IGF2 gene or H19 gene that includes the DMR. For example, primers can be selected
which bind to regions within SEQ ID NOs:35 or 36 outside of the DMR region, but which
can be used to analyze all or a portion of a DMR. Accordingly, certain aspects of the
present invention analyze a DMR that consists essentially of SEQ ID NO:1 or SEQ ID

NO:6, which indicates that a region of the IGF2 or H19 genes, respectively, is analyzed, that includes SEQ ID NO:1 or SEQ ID NO:6, but that also includes other sequences that flank SEQ ID NO:1 or SEQ ID NO:6. For example, 25, 50, 100, 200, or 250 additional flanking nucleotides can be included in the analysis of the DMRs of SEQ ID NO:1 or SEQ ID NO:6.

[0042] Hypomethylation of a DMR is present when there is a measurable decrease in methylation of the DMR. Methods for determining methylation states are provided herein. For example, the H19 DMR can be determined to be hypomethylated when it is methylated at less than 10, less than 5, or less than 3 sites of all of the greater than 25 methylation sites within the H19 DMR. Alternatively, as illustrated in the Examples provided herein, hypomethylation of the H19 DMR can be identified when less than 50% or less than 75% of the methylation sites analyzed are not methylated.

[0043] A fragment or portion of a H19 or IGF2 DMR is a fragment or portion of SEQ ID NO:6 or SEQ ID NO:1, respectively, that for the IGF2 DMR includes at least one methylation site, and preferably includes at least 2, and more preferably at least 3 methylation sites. In certain examples of the H19 and the IGF2 DMR, the fragment includes at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or 25 methylation sites. Methylation sites correspond to positions that are methylated *in vivo* in a normally methylated, aberrantly methylated, or hypermethylated state. The fragment includes a sufficient number of methylation sites to provide information regarding the methylation status of the entire H19 or IGF2 DMR or an important region of the H19 or IGF2 DMR. (See e.g., Nakagawa et al., *Proc. Natl. Acad. Sci.*, 98:591-96 (2001), incorporated by reference in its entirety). As illustrated in Example 2, many of the sites of methylation with the IGF2 DMR and H19 DMR are typically methylated in a coordinated manner. Therefore, methylation state can be analyzed for these DMRs by analyzing less than all of the methylation sites within the DMR. In certain aspects, the methylation sites are those sites for IGF2 that are located within the fragments amplified by the nested primer pairs SEQ ID NO:2 and SEQ ID NO:3 followed by SEQ ID NO:4 and SEQ ID NO:5, or SEQ ID NO:27 and SEQ ID NO:28 followed by SEQ ID NO:29 and SEQ ID NO:30. For H19, in certain aspects methylation sites of fragments of the present invention are those

found within nested primer pairs SEQ ID NO:23 and SEQ ID NO:24 followed by SEQ ID
NO:25 and SEQ ID NO:26.

[0044]   A fragment of the H19 DMR or IGF2 DMR can be the region of the H19 DMR
or IGF2 DMR that is amplified and/or flanked by primers that correspond to SEQ ID
NOS:2-34. For example, the fragment of the H19 DMR can be the region of the H19 DMR
that is amplified by the primer pair recited in SEQ ID NOS:23 and 24, or the primer pair
recited in SEQ ID NOS:25 and 26, or by the nesting of SEQ ID NOS:23 and 24 followed by
SEQ ID NOS:25 and 26. As another example, the fragment of the IGF2 DMR can be the
region of the IGF2 DMR that is amplified by the primer pair recited in SEQ ID NOS: 27
and 28, or the primer pair recited in SEQ ID NOS:29 and 30, or by the nesting of SEQ ID
NOS:27 and 28 followed by SEQ ID NOS:29 and 30. As another example, the fragment of
the IGF2 DMR can be the region of the IGF2 DMR that is amplified by the primer pair
recited in SEQ ID NOS: 2 and 3, or the primer pair recited in SEQ ID NOs:4 and 5, or by
the nesting of SEQ ID NOS:2 and 3 followed by SEQ ID NOs:4 and 5. In certain aspects
the present invention includes analyzing the biological sample for hypomethylation of
positions within the region of the H19 DMR that are analyzed using the nested primer pairs
SEQ ID NOs:23 and 24, followed by SEQ ID NOs:25 and 26, as illustrated in the Examples
herein. Furthermore, in certain aspects, hypomethylation is analyzed in a DNA region
corresponding to SEQ ID NO:1, an IGF2 DMR. In certain aspects the method includes
analyzing the biological sample for hypomethylation of positions within the region of the
IGF2 DMR that are analyzed using the nested primer pairs SEQ ID NOs:2 and 3, followed
by SEQ ID NOs:4 and 5, or the region analyzed using primer pairs SEQ ID NOs: 27 and 28,
followed by SEQ ID NOs:29 and 30, as illustrated in the Examples herein.

[0045]   The fragment of an H19 DMR or IGF2 DMR in certain aspects, is 25, 50, 100,
150, 200, 250, 500, 1000, 2000, 3000, 4000, 50000, or 60000 nucleotides in length. In
certain aspects, methods of the invention analyze methylation of a H19 DMR or IGF2 DMR
sequence that binds under stringent conditions to the H19 DMR (SEQ ID NO:6) or the
IGF2 DMR (SEQ ID NO:1). Furthermore, in certain aspects, methods of the invention
analyze methylation of an H19 DMR or IGF2 DMR, or a fragment thereof, that is at least

75%, 80%, 85%, 90%, 95%,. 96%, 97%, 98%, 99%, or 99.9% identical to SEQ ID NO:6 or SEQ ID NO:1, respectively, or a portion of SEQ ID NO:6 or SEQ ID NO:1.

[0046]   In certain aspects, the H19 DMR or fragment thereof, includes at least one CTCF binding site, for example CTCF binding site 1 or CTCF binding site 6 (corresponding to nucleotides 7855-8192 of Genbank accession no. AF125183, incorporated herein in its entirety by reference) (SEQ ID NO:36) (See Bell, et al., *Nature* 405:482-85 (2000) (incorporated herein by reference); Hark et al., *Nature* 405:486-89 (2000) (incorporated herein by reference); and Bell et al., *Cell* 98, 387-96 (1999) (incorporated herein by reference)). CTCF binding sites can be identified by aligning sequences of the IGF2 or H19 DMR to known CTCF binding sites (See e.g., Bell et al. 1999, supra, at Figure 6). In one example, a fragment of an H19 DMR includes a fragment of the H19 DMR that includes the canonical CTCF binding sites, nucleotides 3010 to 3172 of SEQ ID NO:6, which corresponds to a portion of the H19 DMR from –5315 to –5153 bp upstream of the start site of transcription of H19 (Cui et al., *Canc. Res.*, 61:4947-4950 (2001), incorporated herein in its entirety by reference). One of the effects of methylation of the DMR upstream of the H19 gene, also referred to herein as the H19 DMR, that has been reported is the abrogation of binding of the transcription factor CTCF (Kanduri et al., *Curr. Biol.*, 10:853-56 (2000); Szabo et al., *Curr. Biol.*, 10:607-10 (2000); Hark et al., *Nature*, 405:486-89 (2000); and Bell et al., *Nature*, 405:482-85 (2000)). CTCF binding can discriminate differentially methylated DMRs on the paternal versus maternal alleles *in vivo* (Kanduri et al., *Curr. Bio.*, 10: 853-856 (2000)).

[0047]   A fragment of an H19 DMR, in certain aspects includes less than all of the repeats found in the H19 gene (Nakagawa et al., 2001, supra). The H19 DMR includes 2 copies of a repeat unit, which includes one 450-bp repeat and seven 400-bp repeats (Id.). Therefore, in one example, a fragment of an H19 DMR includes one copy of the H19 DMR repeat unit. In another example, a fragment of an H19 DMR includes at least one copy of the 450-bp repeat and at least one copy of the 400 base pair repeat (Id.).

[0048]   Polymorphisms of the H19 DMR and the IGF2 DMR have been identified. For example, in the H19 DMR, a polymorphism has been identified at nucleotide 8008 (C/A),

nucleotide *8097 (G/A), and nucleotide 8217 (C/G) of accession no. AF125183 (SEQ ID NO:36) (Nakagawa et al., Proc. Nat. Acad. Sci.,* 98:591-96 (2001), incorporated herein by reference). Furthermore, the annotations of Genbank AF125183 (SEQ ID NO:36) identify nucleotides 4489, 4593, 6368, 7342, 7523, 7547, 7591, 7966, 8008, 8097, 8217, 8271, 9554, and 9839 as sites of nucleotide variations. A skilled artisan will recognize that polymorphic versions of the IGF2 DMR or the H19 DMR, or polymorphic versions of fragments of these DMRs can be analyzed in the methods of the present invention. Polymorphisms or mutations in the H19 DMR or IGF2 DMR can be identified using methods known in the art.

[0049]   Although many conventional genetic mutations have been observed in human cancer, most do not occur at high frequency in the general population. Certain embodiments of the present invention are based on the finding of an association between loss of imprinting (LOI) of the IGF2 gene and family history of colorectal cancer (CRC) and between LOI of the IGF2 gene and present or past personal history of colorectal neoplasia. Accordingly, methods of the present invention analyze common molecular markers of cancer risk to identify an increased risk of developing cancer in a subject. In certain embodiments, the method is a DNA-based blood test for the general population.

[0050]   Certain embodiments of the present invention are based on the finding that loss of imprinting of the IGF2 gene is associated with cancers such as colorectal cancer, and that loss of imprinting of the IGF2 gene is correlated with hypomethylation of both the IGF2 gene and the H19 gene. Accordingly, methods of certain embodiments of the present invention analyze hypomethylation in the IGF2 gene and/or the H19 gene to identify loss of imprinting of the IGF2 gene and to identify an increased risk of developing cancer.

[0051]   Accordingly, the present invention relates to a method for identifying an increased risk of developing cancer in a subject. The method includes analyzing a first biological sample from the subject for loss of imprinting of the IGF2 gene and/or the H19 gene. According to the method, a loss of imprinting is indicative of an increased risk of developing cancer.

[0052]   A method of the present invention can also be used to infer a cancer risk of a subject. As discussed above, the method can include analyzing a first biological sample from the subject for loss of imprinting of the IGF2 gene. According to the method, imprinting status is associated with a cancer risk, for example in certain embodiments a loss of imprinting is indicative of an increased risk of developing cancer.

[0053]   A method of the invention can include analyzing genomic DNA for altered methylation of the IGF2 gene or the H19 gene. The method for example, includes analyzing genomic DNA from the sample for hypomethylation of the IGF2 gene or the H19 gene, wherein hypomethylation is associated with an increased risk of developing cancer. Additionally, the altered methylation can occur upstream of a coding region of a gene, or within either exons or introns. For example, hypomethylation of a DMR in the second intron of the IGF2 gene, or of a DMR upstream of the H19 gene, can be associated with cancer.

[0054]   As illustrated in the Example section, especially Example 1, the present invention in certain embodiments, provides a prognostic test for cancer risk, especially colorectal cancer risk. The population frequency of the hypomethylation of IGF2 is approximately 8% when the sample is a blood sample. In embodiments involving a second sample isolated from colorectal tissue, an additional 10% of the population are positive. Therefore, the present invention provides methods that identify cancer risk at high frequency in the general population. A positive blood test confers an increased risk of colorectal cancer, and potentially can be used to identify high risk patients in the general population, for increased cancer surveillance. The method provides an additional advantage in that a negative test excludes patients from repeat colonoscopic examination who may have a positive family history. Furthermore, the test can be performed on RNA or DNA samples.

[0055]   Loss of imprinting, an epigenetic alteration affecting the insulin-like growth factor II gene (IGF2), is found in normal colonic mucosa of approximately 30% of colorectal cancer (CRC) patients, compared to 10% of those without colorectal neoplasia (Cui, H., et al., *Nat. Med.* 4, 1276-1280 (1998)). Therefore, LOI occurs at a relatively high rate in CRC patients and in patients without colorectal neoplasia. Before the

present invention, however, it was not known, whether LOI in normal cells identifies patients with a history of, presence of, or positive family history for colorectal neoplasia in a population not selected for colorectal cancer.

[0056]   In the study provided in Example 1, 11 of 123 (9.0%) of patients with no family history of CRC showed LOI in lymphocytes, compared to 13 of 49 (27%) with a positive family history (adjusted odds ratio 4.41, 95% CI 1.62-12.0, p = 0.004). Similarly, 7 of 106 (6.6%) patients without past or present colonic neoplasia showed LOI, compared to 12 of 56 (21%) patients with adenomas, and 5 of 9 (56%) patients with CRC (adjusted odds ratios 4.10 [95% CI 1.30-12.8, p = 0.016] and 34.4 [95% CI 6.10-194, p < 0.001], respectively). These data support the usefulness and effectiveness of methods of the present invention in identifying an increased risk of developing cancer.

[0057]   For a blood test of risk assessment to be most practical, a DNA rather than RNA-based test can be used. Accordingly, in certain embodiments, methods of the present invention include analyzing the genomic DNA for hypomethylation of a differentially methylated region (DMR) of IGF2 corresponding to SEQ ID NO:1 and/or a polymorphism of a region of a genome corresponding to SEQ ID NO:1, or a fragment of SEQ ID NO:1 or a polymorphism thereof. As illustrated herein, the vast majority of tissues with LOI (i.e., 11 of 12 tissues analyzed herein) show hypomethylation of IGF2, whereas tissues with normal imprinting show normal methylation of IGF2. Thus, LOI in lymphocytes is linked to hypomethylation of a differentially methylated region of IGF2.

[0058]   Since the present specification discloses that hypomethylation of H19 is associated with loss of imprinting of IGF2, in certain aspects, methods of the present invention include analyzing the genomic DNA for hypomethylation of a differentially methylated region (DMR) of H19, such as the DMR corresponding to SEQ ID NO:6, a polymorphism of a region of a genome corresponding to SEQ ID NO:6, or a fragment of SEQ ID NO:6 or a polymorphism thereof. As illustrated in Example 2, the majority of tissues with LOI show hypomethylation of H19, whereas tissues with normal imprinting, in general show normal methylation of H19. Thus, LOI in lymphocytes is correlated with hypomethylation of a differentially methylated region of H19.

[0059]   A method according to the present invention can be performed during routine clinical care, for example as part of a general regular checkup, on a subject having no apparent or suspected neoplasm such as cancer. Therefore, the present invention in certain embodiments, provides a screening method for the general population. The methods of the present invention can be performed at a younger age than present cancer screening assays, for example where the method can be performed on a subject under 65, 55, 50, 40, 35, 30, 25, or 20 years of age.

[0060]   If the biological sample of the subject in question is found to exhibit LOI, for example as the result of hypomethylation of the DMR of IGF2 corresponding to the polynucleotide of SEQ ID NO:1 or hypomethylation of the DMR of H19 corresponding to SEQ ID NO:6, then that subject is identified as having an increased probability of having cancer. In these embodiments, further diagnostic tests may be carried out to probe for the possibility of cancer being present in the subject. Examples of such further diagnostic tests include, but are not limited to, chest X-ray, carcinoembryonic antigen (CEA) or prostate specific antigen (PSA) level determination, colorectal examination, endoscopic examination, MRI, CAT scanning, or other imaging such as gallium scanning, and barium imaging. Furthermore, the method of the invention can be coincident with routine sigmoidoscopy/colonoscopy of the subject. The method could involve use of a very thin tube, or a digital exam to obtain a colorectal sample.

[0061]   The method of the present invention, especially when used to detect local LOI, can be repeated at regular intervals. While not wanting to be limited to a particular theory, methods directed to detecting local LOI by analyzing a blood sample for LOI, typically identify germline mutations. Therefore, typically one test is sufficient. However, for methods used to detect local LOI, a third sample can be isolated, for example from colorectal tissue, for example at least 2 months after isolation of the second sample For example, the third sample can be isolated at about 1 year after he second sample was isolated. In fact, the method can be repeated annually, for example at an annual routine physical exam. Using this regular testing, a method of the present invention is used to screen for an increased risk of developing colorectal cancer by a method that includes

analyzing the third sample from the subject for loss of imprinting of the IGF2 gene or the H19 gene.

[0062]   Additional diagnostic tests can be performed in the future, even if no cancer is present at the time LOI is detected. For example, if LOI is detected in a biological sample of a subject and indicates an increased risk of contracting cancer, periodic (e.g., every 1 to 12 months) chest X-rays, colorectal examinations, endoscopic examination, MRI, CAT scanning, other imaging such as gallium scanning, and/or barium imaging can be scheduled for that subject. Therefore, in these embodiments, LOI is used as a screening assay to identify subjects for whom more frequent monitoring is justified.

[0063]   The biological sample can be virtually any biological sample, particularly a sample that contains RNA or DNA from the subject. The biological sample can be a tissue sample which contains 1 to 10,000,000, 1000 to 10,000,000, or 1,000,000 to 10,000,000 somatic cells. However, it is possible to obtain samples that contain smaller numbers of cells, even a single cell in embodiments that utilize an amplification protocol such as PCR. The sample need not contain any intact cells, so long as it contains sufficient biological material (e.g., protein or genetic material, such as RNA or DNA) to assess the presence or absence of LOI of IGF2 or H19, such as LOI of the IGF2 gene caused by hypomethylation of IGF2 or H19 in the subject.

[0064]   According to the present invention, the biological or tissue sample can be drawn from any tissue that is susceptible to cancer. For example, the tissue may be obtained by surgery, biopsy, swab, stool, or other collection method. The biological sample for methods of the present invention can be, for example, a sample from colorectal tissue, or in certain embodiments, can be a blood sample, or a fraction of a blood sample such as a peripheral blood lymphocyte (PBL) fraction. Methods for isolating PBLs from whole blood are well known in the art. An example of such a method is provided in the Example section herein. In addition, it is possible to use a blood sample and enrich the small amount of circulating cells from a tissue of interest, e.g., colon, breast, etc. using a method known in the art.

[0065] When the method of the present invention provides a method for identifying an increased risk of developing colorectal cancer, a biological sample can be isolated from the colon. Such a tissue sample can be obtained by any of the above described methods, or by the use of a swab or biopsy. In the case of stomach and esophageal cancers, the tissue sample may be obtained by endoscopic biopsy or aspiration, or stool sample or saliva sample. In the case of leukemia, the tissue sample is typically a blood sample.

[0066] As disclosed above, the biological sample can be a blood sample. The blood sample can be obtained using methods known in the art, such as finger prick or phlebotomy. Suitably, the blood sample is approximately 0.1 to 20 ml, or alternatively approximately 1 to 15 ml with the volume of blood being approximately 10 ml.

[0067] Accordingly, in one embodiment, the identified cancer risk is for colorectal cancer, and the biological sample is a tissue sample obtained from the colon, blood, or a stool sample. In another embodiment, the identified cancer risk is for stomach cancer or esophageal cancer, and the tissue may be obtained by endoscopic biopsy or aspiration, or stool sample or saliva sample. In another embodiment, the identified cancer risk is esophageal cancer, and the tissue is obtained by endoscopic biopsy, aspiration, or oral or saliva sample. In another embodiment, the identified cancer risk is leukemia/lymphoma and the tissue sample is blood.

[0068] In the present invention, the subject is typically a human but also can be any mammalian organism, including, but not limited to, a dog, cat, rabbit, cow, bird, rat, horse, pig, or monkey.

[0069] As mentioned above, for certain embodiments of the present invention, the method is performed as part of a regular checkup. Therefore, for these methods the subject has not been diagnosed with cancer, and typically for these present embodiments it is not known that a subject has a hyperproliferative disorder, such as a colorectal neoplasm.

[0070] Methods of the present invention identify a risk of developing cancer for a subject. A cancer can include, but is not limited to, colorectal cancer, esophageal cancer, stomach cancer, leukemia/lymphoma, lung cancer, prostate cancer, uterine cancer, breast

cancer, skin cancer, endocrine cancer, urinary cancer, pancreas cancer, other gastrointestinal cancer, ovarian cancer, cervical cancer, head cancer, neck cancer, and adenomas. In one aspect, the cancer is colorectal cancer.

[0071]   A hyperproliferative disorder includes, but is not limited to, neoplasms located in the following: abdomen, bone, breast, digestive system, liver, pancreas, peritoneum, endocrine glands (adrenal, parathyroid, pituitary, testicles, ovary, thymus, thyroid), eye, head and neck, nervous (central and peripheral), lymphatic system, pelvic, skin, soft tissue, spleen, thoracic, and urogenital. Typically, as used herein, the hyperproliferative disorder is a cancer. In certain aspects, the hyperproliferative disorder is colorectal cancer.

[0072]   The method can further include analysis of a second biological sample from the subject at a target tissue for loss of imprinting of the IGF2 gene, wherein a loss of imprinting in the second sample is indicative of an increased risk of developing cancer in the target tissue. In certain embodiments, the second biological sample is not a blood sample. For example, the first biological sample can be a blood sample and the second biological sample can be isolated from colorectal tissue. In these embodiments analysis of the blood sample can be performed to identify overall risk of developing cancer, whereas the colorectal sample can be analyzed to identify subjects that have an increased risk of developing colorectal cancer. In certain embodiment, the DMRs of both IGF2 and H19 are analyzed for hypomethylation, wherein hypomethylation of one o r both of the DMRs is associated with an increased risk of developing cancer.

[0073]   In another embodiment, the present invention provides a method for managing health of a subject. The method includes performing the method for identifying an increased risk of developing cancer discussed above and performing a traditional cancer detection method. For example a traditional cancer detection method can be performed if the method for identifying cancer risk indicates that the subject is at an increased risk for developing cancer. Many traditional cancer detection methods are known and can be included in this aspect of the invention. The traditional cancer detection method can include, for example, one or more of chest X-ray, carcinoembryonic antigen (CEA) level determination, colorectal examination, endoscopic examination, MRI, CAT scanning, or

other imaging such as gallium scanning, and barium imaging, and sigmoidoscopy/colonoscopy, a breast exam, or a prostate specific antigen (PSA) assay.

[0074]   In another embodiment, the present invention provides a method for prognosing cancer risk of a subject. The method includes analyzing a first biological sample from the subject for altered methylation of the IGF2 gene or the H19 gene. In certain aspects of the invention, the altered methylation is hypomethylation of SEQ ID NO:1, or a polymorphism and/or fragment thereof, or hypomethylation of SEQ ID NO:6, or a polymorphism and/or fragment thereof. Hypomethylation of either or both genes is indicative of an increased risk of developing cancer. In this aspect of the invention, the first biological sample is typically a blood sample.

[0075]   In another aspect, the present invention provides a method for identifying predisposition to colorectal cancer of a subject. The method includes identifying a loss of imprinting in a biological sample from the subject and correlating the loss with a predisposition to colorectal cancer. Loss of imprinting is associated with an increased predisposition to colorectal cancer. The method includes analyzing a first biological sample from the subject for hypomethylation of a differentially methylated region DMR of IGF2 corresponding to SEQ ID NO:1, or a polymorphism or fragment thereof, or hypomethylation of a DMR of H19 corresponding to SEQ ID NO:6, or a polymorphism or fragment thereof. Hypomethylation of either or both of these DMRs is indicative of an increased risk of developing cancer. In this aspect of the invention, the first biological sample is typically a colorectal sample.

[0076]   In another embodiment, the present invention provides to a method for screening a subject for cancer. The method includes analyzing a first biological sample from the subject for loss of imprinting of the IGF2 gene. According to the method, a loss of imprinting is indicative of an increased risk of developing cancer. The method can include analyzing genomic DNA from the sample for hypomethylation of the IGF2 gene or the H19 gene.

[0077] As disclosed herein, methods of the present invention involve analyzing a biological sample for loss of imprinting of IGF2 gene. Genomic imprinting is an epigenetic modification of a specific parental chromosome in the gamete or zygote that leads to monoallelic or differential expression of the two alleles of a gene in somatic cells of the offspring. Imprinting affects various essential cellular and developmental processes, including intercellular signaling, RNA processing, cell cycle control, and promotion or inhibition of cellular division and growth.

[0078] Genomic imprinting is a parent of origin-specific gene silencing that is epigenetic in origin, i.e. not involving the DNA sequence per se but methylation and likely other modifications heritable during cell division (Feinberg, A.P., in *The Metabolic and Molecular Bases of Inherited Disease*, C. R. Scriver, et al., Eds. (McGraw-Hill, New York, 2002)). Loss of imprinting (LOI) of IGF2 was first discovered in embryonal tumors of childhood, such as Wilms tumor (WT), but is one of the most common alterations in cancer, including ovarian, lung, liver, and colon (Feinberg, A.P., in *The Metabolic and Molecular Bases of Inherited Disease*, C. R. Scriver, et al., Eds. (McGraw-Hill, New York, 2002)). The consequence of LOI is best understood in WT. Here it serves as a gatekeeper in about half of tumors, especially those that occur with relatively late onset, and leads to increased expression of IGF2 (Ravenel, J.D., et al., *J Natl. Cancer Inst.* **93**, 1698-1703 (2001)), an important autocrine growth factor in a wide variety of cancers including CRC (Lahm, H., et al., *Br. J. Cancer* **65**, 341-346 (1992); M. C. Gelato and J. Vassalotti, *J. Clin. Endocrinol. Metab.* **71**, 1168-1174 (1990); El-Badry, O. M., et al., *Cell Growth Diff.* **1**, 325-331 (1990); Yee, D., et al., *Cancer Res.* **48**, 6691-6696 (1988); Lamonerie, T., et al., *Int. J. Cancer* **61**, 587-592 (1995); and Pommier, G. J., et al., *Cancer Res.* **52**, 3182-3188 (1992)).

[0079] Loss of imprinting can be caused by hypomethylation or hypermethylation of a gene. The present invention includes methods wherein loss of imprinting is identified by hypomethylation of the IGF2 gene or the H19 gene. For example, the loss of imprinting can be the result of hypomethylation of a DMR within the IGF2 gene, corresponding to SEQ ID NO:1, or a polymorphism and/or fragment thereof, particularly positions 87, 90, and 106

within SEQ ID NO:1. or hypomethylation of a DMR within the H19 gene, or a polymorphism and/or fragment thereof, corresponding to SEQ ID NO:6.

[0080] Methods for detecting loss of imprinting are typically quantitative methods for analyzing imprinting status. The presence or absence of LOI may be detected by examining any condition, state, or phenomenon which causes LOI or is the result of LOI. Such conditions, states, and phenomena include, but are not limited to:

1. Causes of LOI, such as the state or condition of the cellular machinery for DNA methylation, the state of the imprinting control region on chromosome 11, the presence of trans-acting modifiers of imprinting, the degree or presence of histone deacetylation;

2. State of the genomic DNA associated with the genes or gene for which LOI is being assessed, such as the degree of DNA methylation;

3. Effects of LOI, such as:

a. Relative transcription of the two alleles of the genes or gene for which LOI is being assessed;

b. Post-transcriptional effects associated with the differential expression of the two alleles of the genes or gene for which LOI is being assessed;

c. Relative translation of the two alleles of the genes or gene for which LOI is being assessed;

d. Post-translational effects associated with the differential expression of the two alleles of the genes or gene for which LOI is being assessed;

e. Other downstream effects of LOI, such as altered gene expression measured at the RNA level, at the splicing level, or at the protein level or post-translational level (i.e., measure one or more of these properties of an imprinted gene's manifestation into various macromolecules); changes in function that could involve, for example, cell cycle, signal transduction, ion channels, membrane potential, cell division, or others (i.e., measure the biological consequences of a specific imprinted gene being normally or not normally imprinted (for example, QT interval of the heart). Another group of macromolecular changes include processes associated with LOI such as histone acetylation, histone deacetylation, or RNA splicing.

[0081] When detecting the presence or absence of LOI by relying on any one of these conditions, states, or phenomena, it is possible to use a number of different specific analytical techniques. In particular, it is possible to use any of the methods for determining the pattern of imprinting known in the art. It is recognized that the methods may vary depending on the gene to be analyzed.

[0082] Conditions, states, and phenomena which may cause LOI and may be examined to assess the presence or absence of LOI include the state or condition of the cellular machinery for DNA methylation, the state of the imprinting control region on chromosome 11, the presence of trans-acting modifiers of imprinting, the degree or presence of histone deacetylation or histone deacetylation, imprinting control center, transacting modulatory factors, changes in chromatin caused by polycomb-like proteins, trithorax-like proteins, human homologues of other chromatin-affecting proteins in other species such as Su(var) proteins in Drosophila, SIR proteins in yeast, mating type silencing in yeast, or XIST-like genes in mammals.

[0083] It is also possible to detect LOI by examining the DNA associated with the gene or genes for which the presence or absence of LOI is being assessed. By the term "the DNA associated with the gene or genes for which the presence or absence of LOI is being assessed" it is meant the gene, the DNA near the gene, or the DNA at some distance from the gene (as much as a megabase or more away, e.g., methylation changes can be that far away, since they act on chromatin over long distances). Typically, for the present invention LOI is identified or analyzed or detected by detecting hypomethylation of a DMR of the IGF2 gene and/or of a DMR of the H19 gene, as described herein.

[0084] The degree of methylation in the DNA, associated with the gene or genes for which the presence or absence of LOI is being assessed, can be measured or identified using a number of analytical techniques. As discussed above, the method in certain aspects of the invention, detects LOI by detecting hypomethylation of a DMR of the IGF2 gene, which corresponds to SEQ ID NO:1, or a polymorphism and/or fragment thereof. In certain aspects of the invention also discussed above, the method detects LOI by detecting

hypomethylation of a DMR of the H19 gene, which corresponds to SEQ ID NO:6, or a polymorphism and/or fragment of the H19 DMR.

[0085] Numerous methods for analyzing methylation status of a gene are known in the art and can be used in the methods of the present invention to identify either hypomethylation or hypermethylation of the H19 gene or the IGF2 gene. As illustrated in the Examples herein, analysis of methylation can be performed by bisulfite genomic sequencing. Accordingly, denatured genomic DNA can be treated with freshly prepared bisulfite solution at 55°C in the dark overnight, followed by column purification and NaOH treatment, as described in more detail in the Examples section. Bisulfite treatment modifies DNA converting unmethylated, but not methylated, cytosines to uracil. In aspects of the invention in which the IGF2 DMR is analyzed, treated DNA can be amplified by PCR, using primers 5'-GGTGAGGATGGGTTTTTGTT-3' (SEQ ID NO:2) and 5'-CTACTCTCCCAACCTCCCTAA-3' (SEQ ID NO:3), annealing at 55 °C, followed by nested PCR using primers 5'-ATTGGGGGTGGAGGGTGTAT-3' (SEQ ID NO:4) and 5'-TCTATTACACCCTAAACCCAA-3' (SEQ ID NO:5), annealing at 52 °C. Other conditions can include those described previously (Cui, H., et al., *Nat. Med.* 4, 1276-1280 (1998)).

[0086] In another aspect, the IGF2 DMR sequence analyzed corresponds to GenBank nucleotides 631-859 (accession no. Y13633) and can be analyzed after bisulfite treatment using primers 5'-GGGAATGTTTATTTATGTATGAAG-3' (SEQ ID NO:27) and 5'-TAAAAACCTCCTCCACCTCC-3' (SEQ ID NO:28), annealing at 55°C. followed by 5'-TAATTTATTTAGGGTGGTGTT-3' (SEQ ID NO:29) and 5'-TCCAAACACCCCCACCTTAA-3' (SEQ ID NO:30), annealing at 50°C. Other conditions can be those described in Cui et al., *Cancer Research*, 61:4847-4950 (2001) (incorporated herein in its entirety by reference).

[0087] In another aspect, the bisulfite genomic sequencing can be performed of the H19 DMR using the following primers after bisulfite treatment:

[0088]   5'-GAGTTTGGGGGTTTTTGTATAGTAT-3' (SEQ ID NO:23) and
5'-CTTAAATCCCAAACCATAACACTA-3' (SEQ ID NO:24), followed by

[0089]   5'-GTATATGGGTATTTTTTGGAGGT-3' (SEQ ID NO:25) and
5'-CCATAACACTAAAACCCTCAA-3' (SEQ ID NO:26), both annealing at 50°C. Other
conditions can be those described in Cui et al., *Cancer Research*, 61:4847-4950 (2001).

[0090]   In another aspect, the bisulfite genomic sequencing can be performed of the H19
DMR using the following primers after bisulfite treatment:
5'-GTATAGGTATTTTTGGAGGTTTTTTA-3' (SEQ ID NO:31)and
5'-CCTAAAATAAATCAAACACATAACCC-3' (SEQ ID NO:32). The second PCR
primers were: 5'-GAGGTTTTTTATTTTAGTTTTGG-3' (SEQ ID NO:33) and
5'-ACTATAATATATAAACCTACAC-3' (SEQ ID NO:34).

[0091]   In another aspect, genomic sequencing can be performed on the H19 DMR using
the following primer pairs:

[0092]   H1 5'-ATCTTGCTGACCTCACCAAGG-3' (SEQ ID NO:7) and
5'-CGATACGAAGACGTGGTGTGG-3' (SEQ ID NO:8);

[0093]   H2 5'-CCGACTAAGGACAGCCCCCAAA-3' (SEQ ID NO:9) and
5'-TGGAAGTCTCTGCTCTCCTGTC-3' (SEQ ID NO:10);

[0094]   H3 5'-ACAGTGTTCCTGGAGTCTCGCT-3' (SEQ ID NO:11)
5'-CACTTCCGATTCCACAGCTACA-3' (SEQ ID NO:12);

[0095]   H4 5'-ACAGGGTCTCTGGCAGGCTCAA-3' (SEQ ID NO:13)
5'-ATGAGTGTCCTATTCCCAGATG-3' (SEQ ID NO:14);

[0096]   H5 5'-AACTGGGGTTCGCCCGTGGAA -3' (SEQ ID NO:15)
5'-CAAATTCACCTCTCCACGTGC-3' (SEQ ID NO:16);

[0097]   H6 5'-GATCCTGATGGGGTTAGGATGT-3' (SEQ ID NO:17)
5'-GGAATTTCCATGGCATGAAAAT-3' (SEQ ID NO:18);

[0098]   H7 5'-GGTCTGCCTTGGTCTCCTAACT-3' (SEQ ID NO:19)

5'-GGCCACTTTCCTGTCTGAAGAC-3' (SEQ ID NO:20); and

[0099]   H8 5'-CAGTCTCCACTCCACTCCCAAC-3' (SEQ ID NO:21)

5'-GACCTCTCCCTCCCAGACCACT-3' (SEQ ID NO:22).

[0100]   It will be recognized for the primers of the invention that depending on the site bound by the primer and the direction of extension from a primer, that the primers listed above can be used in different pairs. Furthermore, it will be recognized that additional primers can be identified within the IGF2 and H19 DMRs, especially primers that allow analysis of the same methylation sites as those analyzed with primers that correspond to SEQ ID NOs:1-5, and 7-34. In certain aspects, primers allow analysis of the same methylation sites as the primer pairs that correspond to SEQ ID NOs:2 and 3, 4 and 5, 23 and 24, 25 and 26, 27 and 28 and the primer pair that corresponds to SEQ ID NOS: 29 and 30. The regions amplified and/or otherwise analyzed using the above primer pairs can be readily identified by a skilled artisan using sequence comparison tools and/or by analyzing nucleotides fragments that are replicated using the primers. Some of the primers are intended for use after bisulfite treatment. These primer pairs include SEQ ID NOS:2 and 3, 4 and 5, 23 and 24, 25 and 26, 27 and 28, 29 and 30, 31 and 32, and the primer pair corresponding to SEQ ID NOs:33 and 34. Therefore, it will be understood that identification of the binding sites for these primers using computational methods, will take into account that the primers can preferably bind to a polynucleotide whose sequence is modified by bisulfite treatment. For example, for IGF2 DMR analysis, the primers that correspond to SEQ ID NOs:27-30 bind to the following positions, based on the Genbank Y13633 sequence, incorporated herein by reference in its entirety (SEQ ID NO:35): F1: 631-654, R1: 840-859, F2: 688-708, R2: 823-842. As another example, for H19 DMR analysis, the primers that correspond to SEQ ID NOs:31-34 (which for example are used to analyze CTCF binding site 1) bind to the following positions, based on the Genbank AF087017 sequence, incorporated herein by reference (SEQ ID NO:37), F1: 2995-3020, R1: 3284-3309, F2: 3010-3031, R2: 3151-3172. As another example, for H19 DMR analysis, the primers that correspond to SEQ ID NOs:23-26 (which for example are used to

analyze CTCF binding site 6) bind to the following positions, based on the Genbank AF087017 sequence, incorporated herein by reference, F1: 6083-6107, R1: 6397-6420, F2: 6104-6125, R2: 6387-6407.

[0101] Bisulfite treatment can be carried out using the CpG Genome DNA Modification kit (Intergen, Purchase, NY) with the following modifications of the manufacturer's protocol: denatured genomic DNA (4 μg) can be incubated at 55°C in the dark overnight in 1100 μl of freshly prepared Reagent I, with subsequent column purification with the QIAquick PCR purification kit (Qiagen). Purified DNA can be treated at 37°C for 15 min with freshly prepared 3 M NaOH to a final concentration of 0.3 M NaOH. Then the DNA can be precipitated with ethanol and dissolved in 40 μl of 10 mM Tris (pH 8)-1 mM EDTA for nested PCR. PCR products were purified on 2% agarose gels for direct sequencing as described above. The annealing temperature was 55°C. For sequencing individual clones, the PCR products can be subcloned into a TA Cloning vector (Invitrogen, Carlsbad, CA) according to the manufacturer's instructions, and a series of clones, such as 10–15 clones, can be selected for sequencing.

[0102] PCR products can be purified using the QIAEX II gel extraction kit (Qiagen) and directly sequenced with an ABI Prism 377 DNA sequencer using the BigDye™ Terminator Cycle Sequencing kit following the manufacturer's protocol (PE Applied Biosystems, Foster City, CA).

[0103] Certain embodiments of the present invention provide the above-listed primer pairs, typically in an isolated form, and/or the above-listed primers. Accordingly, in these embodiments, the present invention provides isolated oligonucleotides and primer pairs corresponding to SEQ ID NOS:2 and SEQ ID NO:3, SEQ ID NO:4 and SEQ ID NO:5, SEQ ID NO:6 and SEQ ID NO:7, SEQ ID NO:8, and SEQ ID NO:9, SEQ ID NO:9 and SEQ ID NO:10, SEQ ID NO:11 and SEQ ID NO:12, SEQ ID NO:13 and SEQ ID NO:14; SEQ ID NO:15 and SEQ ID NO:16, SEQ ID NO:17 and SEQ ID NO:18, SEQ ID NO:19 and SEQ ID NO:20, SEQ ID NO:21 and SEQ ID NO:22, SEQ ID NO:23 and SEQ ID NO:24, SEQ ID NO:25 and SEQ ID NO:26, SEQ ID NO:27 and SEQ ID NO:28, SEQ ID NO:29 and SEQ ID NO:30. In certain aspects the present invention provides isolated primer pairs,

including SEQ ID NO:2 and SEQ ID NO:3; SEQ ID NO:4 and SEQ ID NO:5;
SEQ ID NO:23 and SEQ ID NO:24; SEQ ID NO:25 and SEQ ID NO:26; SEQ ID NO:27
and SEQ ID NO:28; and SEQ ID NO:29 and SEQ ID NO:30.

[0104]   Altered methylation can be identified by identifying a detectable difference in
methylation. For example, hypomethylation can be determined by identifying whether after
bisulfite treatment a uracil or a cytosine is present at residues corresponding to position 87,
90, and 106 of SEQ ID NO:1. If uracil is present after bisulfite treatment, then the residue
is unmethylated. Hypomethylation is present when there is a measurable decrease in
methylation, for example a measurable decrease in methylation of residues corresponding to
positions 87, 90, and 106 of SEQ ID NO:1, as illustrated in the Example section herein, or a
measurable decrease in methylation of residues corresponding to methylated positions
within the polynucleotides analyzed using the primers disclosed herein.

[0105]   In an alternative embodiment, the method for analyzing methylation of the DMR
can include amplification using a primer pair specific for methylated residues within a DMR
of the IGF2 gene or the H19 gene. In these embodiments, selective hybridization or binding
of at least one of the primers is dependent on the methylation state of the target DNA
sequence (Herman et al., *Proc. Natl. Acad. Sci. USA*, 93:9821 (1996)). For example, the
amplification reaction can be preceded by bisulfite treatment, and the primers can
selectively hybridize to target sequences in a manner that is dependent on bisulfite
treatment. For example, one primer can selectively bind to a target sequence only when one
or more base of the target sequence is altered by bisulfite treatment, thereby being specific
for a methylated target sequence.

[0106]   Other methods are known in the art for determining methylation status of a gene,
such as the IGF2 gene or the H19 gene, including, but not limited to, array-based
methylation analysis and Southern blot analysis.

[0107]   Methods using an amplification reaction, for example methods above for
detecting hypomethylation of the IGF2 DMR or the H19 DMR, can utilize a real-time
detection amplification procedure. For example, the method can utilize molecular beacon

technology (Tyagi S., et al., *Nature Biotechnology*, 14: 303 (1996)) or Taqman™ technology (Holland, P. M., et al., *Proc. Natl. Acad. Sci. USA*, 88:7276 (1991)).

[0108]   Also methyl light (Trinh BN, Long TI, Laird PW. DNA methylation analysis by MethyLight technology, Methods, 25(4):456-62 (2001), incorporated herein in its entirety by reference), Methyl Heavy (Epigenomics, Berlin, Germany), or SNuPE (single nucleotide primer extension) (See e.g., Watson D., et al., *Genet Res.* 75(3):269-74 (2000)). Can be used in the methods of the present invention related to identifying altered methylation of IGF2.

[0109]   As used herein, the term "selective hybridization" or "selectively hybridize" refers to hybridization under moderately stringent or highly stringent physiological conditions, which can distinguish related nucleotide sequences from unrelated nucleotide sequences.

[0110]   As known in the art, in nucleic acid hybridization reactions, the conditions used to achieve a particular level of stringency will vary, depending on the nature of the nucleic acids being hybridized. For example, the length, degree of complementarity, nucleotide sequence composition (for example, relative GC:AT content), and nucleic acid type, i.e., whether the oligonucleotide or the target nucleic acid sequence is DNA or RNA, can be considered in selecting hybridization conditions. An additional consideration is whether one of the nucleic acids is immobilized, for example, on a filter. Methods for selecting appropriate stringency conditions can be determined empirically or estimated using various formulas, and are well known in the art (see, for example, Sambrook et al., supra, 1989).

[0111]   An example of progressively higher stringency conditions is as follows: 2X SSC/0.1% SDS at about room temperature (hybridization conditions); 0.2X SSC/0.1% SDS at about room temperature (low stringency conditions); 0.2X SSC/0.1% SDS at about 42°C (moderate stringency conditions); and 0.1X SSC at about 68°C (high stringency conditions). Washing can be carried out using only one of these conditions, for example, high stringency conditions, or each of the conditions can be used, for example, for 10 to 15 minutes each, in the order listed above, repeating any or all of the steps listed.

[0112]   The present invention can include performing more than one assay for detecting LOI of the IGF2 gene or the H19 gene. For example, a method for detecting LOI of the IGF2 gene by detecting hypomethylation of a DMR of the IGF2 gene corresponding to SEQ ID NO:1 can be performed along with detection of hypomethylation of a DMR of the H19 gene corresponding to SEQ ID NO:6, and/or along with methods that analyze expression of alleles that are affected by imprinting to increase the accuracy and/or sensitivity of the assay.

[0113]   Methods of the present invention in certain aspects, involve analyzing genomic DNA for hypomethylation of a core sequence within SEQ ID NO:1 or within SEQ ID NO:6. Sequences present on a genome, typically the human genome, within the portion of the IGF2 gene corresponding to SEQ ID NO:1, likely will show variable alteration, as found near the H19 and other DMRs in development (Davis, T. L., et al., *Hum. Mol. Genet* 9, 2885-2894 (2000)). In view of the present disclosure, an ordinary artisan can use standard techniques to identify a core sequence within SEQ ID NO:1 for hypomethylation. For example, the amplification product of the amplification reaction disclosed above, can be sequenced with and without bisulfite treatment. An analysis of the sequence will reveal the individual residues that are methylated. As another example, a series of primers can be constructed that selectively hybridize to a series of target sequences within SEQ ID NO:1 or SEQ ID NO:6, in a manner that depends on the methylation state of the target sequence before bisulfite treatment.

[0114]   The degree of methylation in the DNA associated with the gene or genes for which the presence or absence of LOI is being assessed, may be measured by fluorescent *in situ* hybridization (FISH) by means of probes which identify and differentiate between genomic DNAs, associated with the gene for which the presence or absence of LOI is being assessed, which exhibit different degrees of DNA methylation. FISH is described in the Human chromosomes: principles and techniques (Editors, Ram S. Verma, Arvind Babu Verma, Ram S.) 2nd ed., New York: McGraw-Hill, 1995, and de Capoa A., Di Leandro M., Grappelli C., Menendez F., Poggesi I., Giancotti P., Marotta, M. R., Spano A., Rocchi M., Archidiacono N., Niveleau A. Computer-assisted analysis of methylation status of

individual interphase nuclei in human cultured cells. *Cytometry*. 31:85-92, 1998 which is incorporated herein by reference. In this case, the biological sample will typically be any which contains sufficient whole cells or nuclei to perform short term culture. Usually, the sample will be a tissue sample that contains 10 to 10,000, or, for example, 100 to 10,000, whole somatic cells.

[0115]   Additionally, as mentioned above, methyl light, methyl heavy, and array-based methylation analysis can be performed, by using bisulfite treated DNA that is then PCR-amplified, against microarrays of oligonucleotide target sequences with the various forms corresponding to unmethylated and methylated DNA.

[0116]   As mentioned above, methods for detecting LOI can identify altered methylation patterns. However, other methods for detecting LOI are known. For example, certain methods for detecting LOI identify allele-specific gene expression and rely upon the differential transcription of the two alleles. For these methods, RNA is reverse transcribed with reverse transcriptase, and then PCR is performed with PCR primers that span a site within an exon where that site is polymorphic (i.e., normally variable in the population), and this analysis is performed on an individual that is heterozygous (i.e., informative) for the polymorphism. A number of detection schemes can be used to determine whether one or both alleles is expressed. See also, Rainier et al. (1993) *Nature* 362:747-749; which teaches the assessment of allele-specific expression of IGF2 and H19 by reverse transcribing RNA and amplifying cDNA by PCR using new primers that permit a single round rather than nested PCR; Matsuoka et al. (1996) *Proc. Natl. Acad Sci USA* 93:3026-3030 which teaches the identification of a transcribed polymorphism in p57^{KIP2}; Thompson et al. (1996) *Cancer Research* 56:5723-5727 which teaches determination of mRNA levels by RPA and RT-PCR analysis of allele-specific expression of p57^{KIP2}; and Lee et al. (1997) *Nature Genet.* 15:181185 which teaches RT-PCR SSCP analysis of two polymorphic sites. Such disclosures are herein incorporated by reference. In this case, the biological sample will be any which contains sufficient RNA to permit amplification and subsequent reverse transcription followed by polymerase chain reaction. Typically, the biological sample will

be a tissue sample which contains 1 to 10,000,000, 1000 to 10,000,000, or 1,000,000 to 10,000,000, somatic cells.

[0117]   Quantitative analysis of IGF2 or H19 imprinting status can be performed by Hot-stop PCR on cDNA (Uejima, H., et al., *Nat. Genet.* **25**, 375-376 (2000)), as illustrated in the Example section herein. LOI index can be calculated by quantitating the PCR product of a less active allele, or a more active allele, X 100%. LOI can be defined as an LOI index greater than 25 (Cui, H., et al., *Nat. Med.* **4**, 1276-1280 (1998)). In certain embodiments, methods of the present invention that analyze LOI are performed in subjects that are informative for either an APA I or CA repeat polymorphism within exon 9 of IGF2 (Cui, H., et al., *Cancer Research* **62**, 6442-6 (2002)).

[0118]   Methods that detect hypomethylation or hypermethylation have the advantage over some other LOI assays in that they are not restricted to subjects with imprinted polymorphisms that have altered transcription levels.

[0119]   It is also possible to utilize allele specific RNA-associated in situ hybridization (ASISH) to detect the presence or absence of LOI by relying upon the differential transcription of the two alleles. In ASISH, the relative abundance of transcribed mRNA for two alleles is assessed by means of probes which identify and differentiate between the mRNA transcribed from the two alleles. Typically, the probes are tagged with fluorescent labels which results in a high sensitivity and easily quantifiable results. ASISH is described in Adam et al. (1996) "Allele-specific in situ hybridization (ASISH) analysis: a novel technique which resolves differential allelic usage of H19 within the same cell lineage during human placental development," *Development* 122:83-47, which is incorporated herein by reference. In this case, the biological sample will typically be any which contains sufficient whole cells or nuclei to perform histological section and in situ hybridization. Usually, the sample will be a tissue sample which contains for example, 10-100,000, or 100-1000, whole somatic cells.

[0120]   According to the present invention, it is also possible to detect LOI by examining allele-specific post-transcriptional effects (i.e., effects after transcription and before

translation), like alternate splicing that depends on which allele was transcribed, and detection of secondary structure of the RNA.

[0121]   It is also possible, according to the present invention, to detect LOI by examining the relative translation of the two alleles of the gene or genes for which the presence or absence of LOI is being measured. In this case, the presence or relative abundance of the two polypeptides arising from the expression of the two alleles is measured directly. This approach can be effected by any known technique for detecting or quantifying the presence of a polypeptide in a biological sample. For example, allele-specific translational effects may be examined by quantifying the proteins expressed by the two alleles using antibodies specific for each allele (transcribed, translated polymorphism). Such effects may be measured and/or detected by such analytical techniques as Western blotting, or use of an ELISA assay. In this case, the biological sample will be any which contains a sufficient amount of the polypeptide(s) encoded by the gene(s) for which the presence or absence of LOI is being measured.

[0122]   LOI may also be detected by examining post-translational effects, such as secondary modifications that are specific to one allele, like glycosylation or phosphorylation. For example, one allele may be modified, say by phosphorylation or glycosylation, and the other one not. Because the polymorphism encodes a recognition motif, then one can readily distinguish the difference by a Western blot, detecting alternate migration of the polypeptide or protein; use of antibodies specific for the modified form; radioactive incorporation of phosphoryl group or glycosyl group or other modification (i.e., in living cells, followed by the detection of a band at a varying location).

[0123]   LOI may also be detected by reliance on other allele-specific downstream effects. For example, depending on the metabolic pathway in which lies the product of the imprinted gene; the difference will be 2X versus 1X (or some number in between) of the product, and therefore the function or a variation in function specific to one of the alleles. For example, for IGF2, increased mitogenic signaling at the IGF1 receptor, increased occupancy of the IGF1 receptor, increased activity at the IGF2 catabolic receptor, decreased apoptosis due to the dose of IGF2; for KvLQT1, change in the length of the QT interval

depending on the amount and isoform of protein, or change in electrical potential, or change in activity when the RNA is extracted and introduced into *Xenopus* oocytes.

[0124]   It is also possible to detect LOI by detecting an associated haplotype, i.e., linked polymorphisms that identify subjects whose genes are prone to LOI.

[0125]   LOI can be detected by relying on a polymorphism, i.e., a genetic difference between the two alleles. However, it will be recognized that many of the techniques described above may be used to detect LOI even when there is no polymorphism in the two alleles of the gene or genes for which the presence or absence of LOI is being measured. For example, LOI may be detected by reliance on allele-specific DNA methylation (polymorphism independent); histone acetylation; other modifications to DNA; or alterations in replication timing, when the imprinted allele shows "replication timing asynchrony" i.e. the two alleles replicate at different times. When the two alleles replicate at the same time, LOI may be detected by FISH. Since imprinted alleles pair in the late S phase, LOI may be detected by the absence of such pairing in the late S as observed by FISH.

[0126]   On the other hand certain techniques are more conveniently used when there is a polymorphism in the two alleles of the gene or genes for which the presence or absence of LOI is being measured. For example, RT-PCR followed by SSCP (single strand conformational polymorphism) analysis; restriction enzyme digestion analysis followed by electrophoresis or Southern hybridization; or radioisotopic PCR; PCR; allele-specific oligonucleotide hybridization; direct sequencing manually or with an automated sequencer; denaturing gradient gel electrophoresis (DGGE); and many other analytical techniques can be used to detect LOI when relying on a polymorphism.

[0127]   In another embodiment of the present invention the method involves measuring the degree of LOI such as by measuring the degree of hypomethylation of a DMR for a particular gene or set of genes. In certain embodiments, the method includes measuring the degree of hypomethylation of the DMR of IGF2 gene corresponding to SEQ ID NO:1 or a polymorphism thereof, or a fragment thereof. In another embodiment, the method includes

measuring the degree of hypomethylation of the DMR of the H19 gene corresponding to SEQ ID NO:6 or a polymorphism thereof, or a fragment thereof.

[0128] As used herein, when hypomethylation is measured, "the degree of LOI" means the percentage of methylation compared to a fully methylated DMR. As used herein, when expression of different polymorphisms is compared, "the degree of LOI" means total expression (as measured by actual expression or transcription) attributable to the allele which is normally imprinted. The degree of LOI may be calculated by allele ratio, i.e., the more abundant allele divided by the less abundant allele. The degree of LOI may be determined by any method which allows the determination of the relative expressions of the two alleles. For example, a degree of LOI of 100% reflects complete LOI (equal expression of both alleles), while a degree of LOI of 0% reflects no LOI (expression of only one allele). Any method of measuring the relative expression of the two alleles is considered to be included in the present invention.

[0129] The degree of LOI can be measured for the IGF2 gene or the H19 gene when screening for the presence of colorectal cancer, or other cancers, e.g., the degree of LOI is measured for the IFG2 gene or the H19 gene when screening for the presence of stomach cancer, esophageal cancer, or leukemia.

[0130] The degree of LOI can be measured by measuring the degree of hypomethylation of the DMR of IGF2 or the DMR of H19 or a fragment thereof, or a polymorphism thereof, in a blood sample, for example a PBL sample, wherein a high degree of hypomethylation is indicative of an increased risk for cancer. For example, a series of genomic clones can be analyzed that are derived from the subject. These clones can be analyzed for hypomethylation of the DMR of IGF2 or the DMR of H19. The degree of hypomethylation can be determined by identifying the methylation frequency of possibly methylated sites. Possible methylated sites are cytosine residues that are typically methylated in a subject, but become unmethylated in certain subjects that are at an increased risk of developing cancer. For example, residues corresponding to positions 87, 90, and 106 of SEQ ID NO:1. If uracil is present after bisulfite treatment, then the residue is unmethylated. Hypomethylation is present when in a series of clones of genomic DNA, there is a

measurable decrease in methylation, such as a measurable decrease in methylation of positions 87, 90, and 106 of SEQ ID NO:1.

[0131]   A linear detection platform can be employed to quantitate LOI. A linear detection platform is a detection platform that allows quantitation because the amount of target present and signal detected are linearly related. In this regard, a PhosphorImager (model 445SI, manufactured by Molecular Dynamics), which detects radioactive emissions directly from a gel, can be used. Other linear detection systems include carefully titrated autoradiography followed by image analysis, beta-emission detection analysis (Betascan). Another linear detection platform is an automated DNA sequencer such as ABI 377 analyzer. Another linear detection platform is an array based system with appropriate software. Another is SNuPE.

[0132]   In addition to measuring the degree of imprinting when an imprinted polymorphism is present in a gene, it is possible to assess the degree of LOI in a particular gene even when an imprinted polymorphism is not present in that gene. For example, imprinting can be assessed by the degree of methylation of CpG islands in or near an imprinted gene (e.g., Barletta, Cancer Research, op. cit). In addition, imprinting can be assessed by changes in DNA replication timing asynchrony, e.g. White L M, Rogan P K, Nicholls R D, Wu B L, Korf B. Knoll J H, Allele-specific replication of 15q11-q13 loci: a diagnostic test for detection of uniparental disomy. *American Journal of Human Genetics*. 59:423-30, 1996.

[0133]   On the other hand, certain techniques are more conveniently used when there is a polymorphism in the two alleles of the gene or genes for which the presence or absence of LOI is being measured. For example, RT-PCR, followed by gel electrophoresis to distinguish length polymorphisms, or RT-PCR followed by restriction enzyme digestion, or by automated DNA sequencing, or by single strand conformational polymorphism (SSCP) analysis, or denaturing gradient gel electrophoresis, etc.; or, completely DNA based methods that exploit, for example DNA methylation, which require no RT step, to convert RNA to cDNA prior to PCR).

[0134] Once the degree of LOI, such as the level of hypomethylation, has been measured for the gene or genes in question, the risk of having cancer is then assessed by comparing the degree of LOI for that gene or genes to a known relationship between the degree of LOI and the probability of the presence of the particular type of cancer or other disease. The relationship between the degree of LOI and the probability of the presence of a particular type of cancer may be determined for any combination of a normally imprinted gene or genes and a particular type of cancer by determining.

[0135] When the degree of LOI is measured, such as the degree of IGF2 hypomethylation, the measured degree of LOI is compared to a known relationship between the degree of LOI and the probability of contracting the particular type of cancer. The relationship between the degree of LOI and the probability of contracting a particular type of cancer may be determined by one of ordinary skill in the art for any combination of a normally imprinted gene or genes and a particular type of cancer by determining the degree of LOI in a statistically meaningful number of tissue samples obtained from patients with cancer, and determining the degree of LOI in a statistically meaningful number of tissue samples obtained from patients without cancer, and then calculating an odds ratio as a function of the degree of LOI.

[0136] It should also be understood that measuring the degree of LOI, can be carried out by comparing the degree of LOI against one or more predetermined threshold values, such that, if the degree of LOI is below a given threshold value, which can be manifested in a regular methylation pattern, then the subject is assigned to a low risk population for having cancer, contracting cancer, and/or having replication error repair defects. Alternatively, the analytical technique may be designed not to yield an explicit numerical value for the degree of LOI, but instead yield only a first type of signal when the degree of LOI is below a threshold value and/or a second type of signal when the degree of LOI is below a threshold value. It is also possible to carry out the present methods by means of a test in which the degree of LOI is signaled by means of a non-numeric spectrum such as a range of colors encountered with litmus paper.

[0137]   Example 2 reveals a possible mechanism wherein hypomethylation of the DMR of IGF2 results in loss of imprinting of the IGF2 gene. By this mechanism, methylation of the DMR of IGF2 recruits repressors to the IGF2 gene. Accordingly, in another embodiment, the present invention provides a method for identifying a repressor of IGF2, or a factor that represses IGF2 imprinting, by identifying a factor that preferentially binds to the methylated versus the non-methylated DMR of IGF2. In this embodiment, virtually any method for identifying factors, especially protein factors, that bind methylated nucleic acids and/or affect gene expression can be included. Many such methods are known in the art.

[0138]   For example, a protein preparation can be contacted with an IGF2 DMR or fragment thereof, and binding of a protein of the protein preparation to the IGF2 DMR can be detected. A protein that binds the IGF2 DMR or fragment thereof, can then be analyzed for the ability to repress expression of IGF2, for example using a co-transfection assay, that includes a polynucleotide that includes the IGF2 DMR or fragment thereof, operably linked to a reporter gene, and a polynucleotide encoding the IGF2 DMR-binding protein. Expression of the reporter gene can be analyzed and compared to control experiments that do not include the IGF2 DMR binding protein. Furthermore, the results can be compared in cells with methylated IGF2 DMR and cells with hypomethylated IGF2 DMR, for example DNA methyltransferase knockout cells (See Example 2). The method provides for the identification of IGF2 repressors that preferentially bind methylated IGF2 DMR.

[0139]   Binding of a protein to the IGF2 DMR or a fragment thereof, can be detected, for example, using an electrophoretic mobility shift assay (EMSA). A protein preparation such as a nuclear extract can be prepared (See e.g., Dignam et al., Nucl. Acids Res. 11;1475-1489 (1983)). EMSAs can be performed as described previously (Latinkic and Lau, J. Biol. Chem. 269:23163-23170 (1994)); Wang et al., Cell 87:697-708 (1996)). In assays where unlabeled oligonucleotides (10-fold or 100-fold molar excess) or antibodies directed against known transcription factors are used, these reagents are typically added to the binding reaction prior to the addition of nuclear extract. The antibodies used in most experiments can be selected based on their ability not to interfere with the factor binding to DNA. Factors that specifically bind the DMR of IGF2 when it is methylated can be identified by

comparing mobility shift patterns using methylated and hypomethylated IGF2 DMR of SEQ ID NO:1, or fragments thereof.

[0140] The term "nucleic acid molecule" is used broadly herein to mean a sequence of deoxyribonucleotides or ribonucleotides that are linked together by a phosphodiester bond. As such, the term "nucleic acid molecule" is meant to include DNA and RNA, which can be single stranded or double stranded, as well as DNA/RNA hybrids. Furthermore, the term "nucleic acid molecule" as used herein includes naturally occurring nucleic acid molecules, which can be isolated from a cell, for example, the IGF-2 gene, as well as synthetic molecules, which can be prepared, for example, by methods of chemical synthesis or by enzymatic methods such as by the polymerase chain reaction (PCR), and, in various embodiments, can contain nucleotide analogs or a backbone bond other than a phosphodiester bond.

[0141] The terms "polynucleotide" and "oligonucleotide" also are used herein to refer to nucleic acid molecules. Although no specific distinction from each other or from "nucleic acid molecule" is intended by the use of these terms, the term "polynucleotide" is used generally in reference to a nucleic acid molecule that encodes a polypeptide, or a peptide portion thereof, whereas the term "oligonucleotide" is used generally in reference to a nucleotide sequence useful as a probe, a PCR primer, an antisense molecule, or the like. Thus, a polynucleotide of the invention can encode, for example SEQ ID NO:1, whereas an oligonucleotide of the invention can be used as a probe to detect SEQ ID NO:1. Of course, it will be recognized that an "oligonucleotide" also can encode a peptide. As such, the different terms are used primarily for convenience of discussion.

[0142] A polynucleotide or oligonucleotide comprising naturally occurring nucleotides and phosphodiester bonds can be chemically synthesized or can be produced using recombinant DNA methods, using an appropriate polynucleotide as a template. In comparison, a polynucleotide comprising nucleotide analogs or covalent bonds other than phosphodiester bonds generally will be chemically synthesized, although an enzyme such as T7 polymerase can incorporate certain types of nucleotide analogs into a polynucleotide

and, therefore, can be used to produce such a polynucleotide recombinantly from an appropriate template

[0143] In another aspect, the present invention includes kits that are useful for carrying out the methods of the present invention. The components contained in the kit depend on a number of factors, including: the condition, state, or phenomenon relied on to detect LOI or measure the degree of LOI, the particular analytical technique used to detect LOI or measure the degree of LOI, and the gene or genes for which LOI is being detected or the degree of LOI is being measured.

[0144] Accordingly, the present invention provides a kit for determining a methylation status of a differentially methylated region (DMR) of IGF2 corresponding to SEQ ID NO:1 or a fragment or a polymorphism thereof, or of a DMR of H19 corresponding to SEQ ID NO:6 or a fragment or a polymorphism thereof. The kit includes an oligonucleotide probe, primer, or primer pair, or combination thereof for carrying out a method for detecting hypomethylation, as discussed above. For example, the probe, primer, or primer pair, can be capable of selectively hybridizing to the DMR either with or without prior bisulfite treatment of the DMR. The kit can further include one or more detectable labels.

[0145] The kit can also include a plurality of oligonucleotide probes, primers, or primer pairs, or combinations thereof, capable of selectively hybridizing to the DMR with or without prior bisulfite treatment of the DMR. The kit can include an oligonucleotide primer pair that hybridizes under stringent conditions to all or a portion of the DMR only after bisulfite treatment. For example, the kit can include the primer pairs of SEQ ID NOS:2-34 The kit can include instructions on using kit components to identify an increased risk of developing cancer. In certain embodiments the instructions relate to subjects of the general population. The kit for example, includes one or both of a primer pair corresponding to the primer pair SEQ ID NO:23 and SEQ ID NO:24 and the primer pair SEQ ID NO: 25 and SEQ ID NO:26. In another aspect, the kit for example, includes one or both of a primer pair corresponding to the primer pair SEQ ID NO:27 and SEQ ID NO:28, and the primer pair SEQ ID NO: 29 and SEQ ID NO:30.

[0146]   When LOI is detected by relying on the degree of methylation of the genomic DNA associated with the gene(s) for which LOI is being detected or the degree of LOI is being measured using FISH, the kit will typically contain one or more probes which can identify a specific imprinted gene or group of genes. Typically, such probes will be nucleic acids or monoclonal antibodies and will be linked to, for example, a fluorescent label.

[0147]   In the case of detecting LOI by relying on the differential rates of transcription of two polymorphic alleles, the kit can include:

(i) means for the amplification of the mRNAs corresponding to the two polymorphic alleles of the gene in question. Examples of such means include suitable DNA primers for the PCR amplification of the mRNAs corresponding to the two polymorphic alleles of the gene in question. Specific examples of such means include any pair of DNA primers which will anneal to and amplify any gene which is normally imprinted and in which a polymorphism is present.

According to the present invention, the kit may further include:

(ii) means for identifying the products of the amplification of the mRNAs corresponding to the two polymorphic alleles of the gene in question. Such means include, but are not limited to, a restriction enzyme which specifically cleaves one of the products of the amplification of the mRNAs corresponding to the two polymorphic alleles of the gene in question. Specific examples of such enzymes include, but are not limited to, Apa I in the case of the IGF2 gene.

[0148]   When the degree of LOI is measured by relying on the differential rates of transcription of two polymorphic alleles, the kit may comprise:

(i) means for the linear amplification of the mRNAs corresponding to the two polymorphic alleles of the gene in question. Examples of such means include a sufficient quantity of suitable DNA primers for the PCR amplification of the mRNAs corresponding to the two polymorphic alleles of the gene in question, such that the PCR amplification may be carried out without exhausting the primers and linear amplification achieved. Specific

examples of such means includes any pair primers for any gene which is normally imprinted.

According to the present invention, the kit can further include:

(ii) means for identifying the products of the amplification of the mRNAs corresponding to the two polymorphic alleles of the gene in question. Such means include a restriction enzyme which specifically cleaves one of the products of the amplification of the mRNAs corresponding to the two polymorphic alleles of the gene in question.

[0149]   When detecting LOI or measuring the degree of LOI by ASISH, the kit will typically contain one or more probes which can identify and distinguish between the RNA associated with the two alleles. Typically, such probes will be nucleic acids that are specific for each allele, and are used either sequentially or together using different fluorescent labels for each allele.

[0150]   When detecting LOI or measuring the degree of LOI by assessing the relative translation of two alleles, the kit may contain antibodies that distinguish the protein product of the two alleles.

[0151]   The following example is intended to illustrate but not limit the invention.

EXAMPLE 1

ASSOCIATION OF LOSS OF IMPRINTING AND COLORECTAL NEOPLASIA

[0152]   This example illustrates that LOI in normal tissue is associated with either a family history or personal history of colorectal neoplasia.

Materials & Methods

[0153]   Study population. Subjects were identified and recruited in the Johns Hopkins Outpatient Endoscopy Clinic and the Johns Hopkins Greenspring Endoscopy Unit. Eligible subject were those individuals who were having a colonoscopy for any medical indication, who were 18 years of age or older and who had physician approval. Written informed

consent was obtained from all subjects. Clinical, demographic, family history and exposure information were obtained with the use of validated questionnaires. The protocol was approved by the Johns Hopkins Joint Committee on Clinical Investigation.

[0154] Collection of research materials. Colon tissues were collected from patients who underwent a colonoscopy. Ninety-seven percent of patients agreed to participate. Colonoscopic examinations were performed with a standard Olympus colonoscope by several endoscopists. Up to 8 mucosal punch biopsies were obtained from proximal and distal colon using routine biopsy forceps. The tissues were immediately frozen in liquid nitrogen and stored at −135 °C. Lymphocytes were separated from blood (20 ml) from every patient with Accuspin tubes (Sigma/Aldrich, St. Louis, Missouri) using Ficoll-Paque Plus (Amersham Pharmacia Biotech, Pisacataway, New Jersey) and centrifuged at 400g at room temperature for 30 minutes. The lymphocyte layer was collected and washed once with PBS. The isolated lymphocyte pellets were immediately stored at −135 °C until the assays were performed.

[0155] DNA and RNA preparation. DNA extraction was performed as previously described (Cui, H., et al., *Nat. Med.* 4, 1276-1280 (1998), incorporated in its entirety herein by reference). RNA was extracted from frozen mucosal punch biopsies and lymphocytes with the RNeasy Mini Kit (Qiagen, Valencia, CA) shortly before RT. Each RNA sample was quantified by spectrophotometry and agarose gel electrophoresis, treated with 10 u RNase inhibitor (Invitrogen, Carlsbad, CA) and used immediately.

[0156] Quantitative analysis of IGF2 imprinting status. Reverse transcription was performed with freshly extracted RNA samples, which had been treated with DNase I to remove any DNA contamination as previously described (Cui, H., et al., *Nat. Med.* 4, 1276-1280 (1998)). Quantitative analysis of IGF2 imprinting status was performed by Hot-stop PCR (Uejima, H., et al., *Nat. Genet.* 25, 375-376 (2000), incorporated herein in its entirety, by reference). All reactions were done in duplicate with presence and absence of reverse transcriptase on identical RNA sample to exclude any possibility of DNA contamination.

[0157]   Detection of DNA Methylation. The IGF2 DMR is located on positions –566 to
–311 relative to IGF2 exon 3 (corresponding to positions 661 to 916 of GenBank accession
no. Y13633). Analysis of methylation was performed by bisulfite genomic sequencing by
analyzing positions 87, 90, and 106 of the IFG2 DMR (SEQ ID NO:1). Genomic DNA was
treated with bisulfite, as follows: Sodium hydroxide was added to genomic DNA to a final
concentration of 0.3M and incubated at 37 °C for 20 min to denature the genomic DNA. A
sodium metabisulfite solution is added to the denatured DNA to a concentration of about
2M sodium metabisulfite (EM Sience) and 10 mM hydroquinone (EM Science), pH 7.0, and
the DNA in the sodium metabisulfite solution is incubated at 55 °C in the dark for about 12
– 16 hrs. Treated DNA was purified with the QIAquick PCR purification kit according
manufacturer's protocol. The purified DNA was exposed to about 1/10 volume of 3M
NaOH and incubated at 37 °C for 20 min. Five molar NH_4OAC to a final 3M concentration
and kept at room temperature for 5 min to neutralize DNA. Five µg yeast RNA and 3X
volume of pure ethanol and were added to the neutralized DNA, and the solutions were kept
at –80 °C for 30 min and then centrifuged to precipitate DNA. The DNA was then washed
with 70% ethanol, dissolved in TE (pH 7.2). and stored at –20 °C for PCR.0

[0158]   PCR was performed as follows: Treated DNA was amplified by PCR, using
primers 5'-GGTGAGGATGGGTTTTTGTT-3' (SEQ ID NO:2) and
5'-CTACTCTCCCAACCTCCCTAA-3' (SEQ ID NO:3), annealing at 55 °C, followed by
nested PCR using primers 5'-ATTGGGGGTGGAGGGTGTAT-3' (SEQ ID NO:4) and
5'-TCTATTACACCCTAAACCCAA-3' (SEQ ID NO:5), annealing at 52 °C.

[0159]   PCR products were purified on 2% agarose gels for direct sequencing as
described above. The annealing temperature was 55°C. For sequencing individual clones,
the PCR products were subcloned into a TA Cloning vector (Invitrogen, Carlsbad, CA)
according to the manufacturer's instructions, and 15-20 clones were selected for
sequencing.

[0160]   All of the PCR products were purified using the QIAEX II gel extraction kit
(Qiagen) and directly sequenced with an ABI Prism 377 DNA sequencer using the BigDye

Terminator Cycle Sequencing kit following the manufacturer's protocol (PE Applied Biosystems, Foster City, CA) (See Cui et al., 61:4947 (2001), incorporated herein in its entirety, by reference.

[0161]   Statistical Analysis. Hypothesis testing was performed with a combination of chi-square and Fisher exact t-tests as appropriate. Multiple logistic regression models were constructed for the association of the independent covariates with LOI, and for the association of LOI with colorectal neoplasia and family history of cancer. Kappa statistics was used to determine agreement beyond chance. Statistical analysis was performed using STATA 7.0 software (Stata Corp.).

Results

[0162]   To ensure complete ascertainment of colorectal neoplasia, a cross-sectional analysis was performed of patients who provided clinical and family history information. Contemporaneous colonoscopic examination was performed with mucosal biopsy. For logistical purposes, patients were chosen that were selected for colonoscopic examination for clinical indications. For this reason, there was modest enrichment for a past history of colonic adenoma or cancer (13%, compared to 10% in the general U.S. population of this age).

[0163]   Four hundred and twenty one patients agreed to participate between 1999 and 2001. 191 patients were informative for either an APA I or CA repeat polymorphism within exon 9 of IGF2 allowing analysis of imprinting status. In addition to both proximal and distal colonic mucosal specimens, PBL for RNA and DNA analysis, a family and personal history of neoplasia, environmental exposures, medications, and dietary information was obtained.

[0164]   First the relationship between LOI and age was examined. It has been suggested previously that altered IGF2 methylation is age-related, suggesting that epigenetic abnormalities are acquired over time (Issa, J. P., et al., *Proc. Natl. Acad. Sci. U.S.A.* 93, 11757-11762 (1996)). However, no relationship between LOI in PBL and age was found (Table 1). There were also no significant differences by sex or race (Table 1). Next the

relationship between LOI and family history was analyzed. The odds of LOI in PBL were 4.4 times greater in patients with a positive family history of CRC compared to their counterparts with a negative family history (p = 0.003; Table 1).

[0165] The relationship between LOI and colorectal neoplasia in the patient was next evaluated. The odds of LOI in PBL were 4.4 times greater in patients with past or present colorectal neoplasia (adenomatous polyps or cancer) than their counterparts without neoplasia (p = 0.002; Table 1), indicating a strong association between LOI and colorectal neoplasia. Even when patients with a positive family history were excluded from the analysis, the odds of LOI in PBL these odds were 4.7 times greater (95% CI 1.29-17.3, p = 0.01).

[0166] The accepted model for colorectal carcinogenesis is that cancers progress from adenomas (E. R. Fearon and B. Vogelstein, *Cell* 61, 759-767 (1990)). Consequently when stratified, the odds of LOI in PBL were 4.1 times greater in patients with past or present adenomas but no CRC, compared to patients with no past or present neoplasia (p = 0.016; Table 1), and they were 34.4-fold greater in patients with past or present CRC than in those without colorectal neoplasia (p < 0.0001; Table 1). These data strongly suggest that LOI is associated with both initiation and progression of colorectal neoplasia.

Table 1. Association of loss of imprinting (LOI) of IGF2 in peripheral blood lymphocytes with family history of colon cancer and with present or past colonic neoplasia in the patient.

| | Imprinting | | Unadjusted odds ratio | | | Adjusted odds ratio* | | |
| | Normal | LOI | P value | OR | 95% CI | P value | OR | 95% CI |
	N (%)	N (%)						
Age (± SD)	58.7 ± 12.8	59.9 ± 9.7	0.64	1.00	0.97-1.04	0.35	1.02	0.97-1.07
Sex								
Women	72 (85.7)	12 (14.3)						
Men	75 (86.2)	12 (13.8)	0.93	0.96	0.41-2.28	0.70	1.21	0.45-3.23
Race								
White	129 (86.0)	21 (14.0)						
Black	18 (85.7)	3 (14.3)	0.97	1.02	0.28-3.78	0.55	0.62	0.13-2.96
Family history								
No	111 (91.0)	11 (9.0)						
Yes	36 (73.4)	13 (26.5)	0.003	3.64	1.5-8.84	0.004	4.41	1.62-12.0
Colonic Neoplasia								
No	99 (93.4)	7 (6.6)						
Yes	48 (73.8)	17 (26.2)	0.001	5.01	1.94-12.89	0.002	4.37	1.74-11.0
Adenomas	44 (78.6)	12 (21.4)	0.008	3.85	1.42-10.5	0.016	4.10	1.30-12.8
CRC	4 (44.4)	5 (55.6)	<0.0001	17.67	3.85-81.0	<0.0001	34.4	6.10-194

*Hypothesis testing performed with multiple logistic regression adjusted for age, gender, race, family history, and physical activity; OR, odds ratio; 95% CI, 95% confidence interval

[0167] The relationship between LOI in PBL and LOI in the colon was determined in patients from whom informative samples sufficient for imprinting analysis could be obtained in both tissues. All of the patients with LOI in PBL also showed LOI in normal colon (Table 2 and data not shown). In the remainder LOI was limited to the colon (Table 2), and in these patients it was present variably in proximal or distal colonic mucosa (data not shown). Thus, LOI either was a generalized defect affecting both blood and colon, or a focal abnormality within one or more samples within the colon. In the latter group, no statistically significant association with family or personal history of colorectal neoplasia was found.

Table 2. Concordance of imprinting status between colonic mucosa and blood*

	Colon	
Blood	Normal	LOI
Normal	123	21
LOI	0	24

*Kappa statistic 88.0%, p value < 0.0001

[0168] It was next determined whether a method of the present invention can be performed using DNA rather than RNA. SEQ ID NO:1 provides a differentially methylated region (DMR) within IGF2 that shows hypomethylation in CRC with LOI (Cui H. et al., *Cancer Res.* **62**, 6442-6446 (2002), incorporated herein in its entirety by reference). In order to determine whether a hypomethylation defect occurs in PBL and colon of patients without known neoplasia, we examined 24 samples, 12 from normal tissues (6 PBL, 6 matched normal colonic mucosa) with normal imprinting, and 12 from normal tissues (6 PBL, 6 matched normal colonic mucosa) with LOI. In all 12 tissues with normal imprinting, IGF2 showed a normal pattern of half-methylation (Fig. 1A). In contrast, in 11 of 12 samples from normal tissue with LOI, IGF2 showed hypomethylation of the IGF2 DMR; in the other sample, IGF2 showed partial methylation of both alleles but was nevertheless abnormal (Fig. 1B). The significance of hypomethylation between normal

52

tissues with and without LOI was p < 0.0001 (Fisher's exact test). In contrast, H19 showed hypomethylation in all cases, regardless of imprinting status (data not shown). Thus, aberrant IGF2 methylation is linked to LOI in normal colon and lymphocytes, just as it is in CRC.

[0169]   In summary, a strong and significant association of LOI with family history, and with present or past personal history of colorectal neoplasia was identified using methods of the present invention. When present in PBL, LOI appears to be a systemic abnormality, since it was always also present in both proximal and distal colon. It cannot be concluded currently that the abnormality is present in the germline, as it is epigenetic and might be acquired postnatally. This abnormality is common, present in 14% of the patients studied, which by design may be mildly enriched for CRC. Nevertheless, a 10% frequency of LOI in PBL in the general population has previously been observed (Cui, H., et al., *Nat. Med.* **4**, 1276-1280 (1998); Sakatani, T., et al., *Biochem. Biophys. Res. Commun.* **283**, 1124-1130 (2001)).

[0170]   This epigenetic abnormality was present at both the RNA and DNA level. Eleven of 12 tissues with LOI showed hypomethylation of IGF2, and all 12 tissues with normal imprinting showed normal methylation of IGF2. The methylation assay may be improved, as the entire IGF2 DMR has not yet been examined by bisulfite sequencing. There may be a critical core sequence involved, with nearby sequences showing variable alteration, as found near the H19 and other DMRs in development (Davis, T. L., et al., *Hum. Mol. Genet* **9**, 2885-2894 (2000)).

[0171]   The odds ratio for colorectal cancer of LOI (34.4) is higher than seen for mutation of the mismatch repair genes in HNPCC (H. T. Lynch and J. F. Lynch, *Semin. Surg. Oncol.* **18**, 305-313 (2000)), which confers an 80% lifetime risk of CRC (H. T. Lynch and A. de la Chapelle, *J. Med. Genet.* **36**, 801-818 (1999)). In contrast, the I1307K mutation of APC confers only a two-fold increased risk of colorectal cancer (Laken S. J., et al., *Nat. Genet.* **17**, 79-83 (1997)). Furthermore, the prevalence of LOI, 10%, is at least 10-fold higher than all known CRC-predisposing genetic mutations in the population combined

(Samowitz, W. S., et al., *Gastroenterology* **121**, 830-838 (2001); and Percesepe, A., et al., *J. Clin. Oncol.* **19**, 3944-3950 (2001)). Consequently, conventional genetic mutation screening for cancer risk has been targeted at defined populations with a strong family history, and not for screening and surveillance in the general population. In contrast an LOI blood test might be of value for population screening.

EXAMPLE 2

LOSS OF IMPRINTING IN COLORECTAL CANCER LINKED TO HYPOMETHYLATION OF H19 AND IGF2

[0172] This example illustrates that loss of imprinting of IGF2 in colorectal cancer is correlated with hypomethylation of the DMR of IGF2, and in at least some colorectal cancer patients with hypomethylation of the DMR of H19 as well. Furthermore, the example reveals a model of gene regulation based on hypomethylation of the DMR of IGF2.

Materials and Methods

[0173] **Bisulfite Sequencing Analysis.** H19 CTCF binding site 1 (CBS1) was analyzed as described earlier (H. Cui et al, Cancer Res., 61: 4947-4950, 2001); CBS6 corresponds to GenBank nucleotides 7855–8192 (accession no. AF125183) and was analyzed after bisulfite treatment using primers 5'-GAGTTTGGGGGGTTTTTGTATAGTAT-3' (SEQ ID NO:23) and 5'-CTTAAATCCCAAACCATAACACTA-3' (SEQ ID NO:24), followed by 5'-GTATATGGGTATTTTTTGGAGGT-3' (SEQ ID NO:25) and 5'-CCATAACACTAAAACCCTCAA-3' (SEQ ID NO:26), both annealing at 55°C.

[0174] The *IGF2* DMR sequence analyzed corresponds to GenBank nucleotides 631–859 (accession no. Y13633), and was analyzed after bisulfite treatment using primers 5'-GGGAATGTTTATTTATGTATGAAG-3' (SEQ ID NO:27) and 5'-TAAAAACCTCCTCCACCTCC-3' (SEQ ID NO:28), annealing at 55°C, followed by 5'-TAATTTATTTAGGGTGGTGTT-3' (SEQ ID NO:29) and

5'-TCCAAACACCCCCACCTTAA-3' (SEQ ID NO:30), annealing at 50°C. Other
conditions are as described earlier (H. Cui et al, Cancer Res., 61: 4947-4950, 2001).

[0175] **Methyltransferase Activity Analysis.** *In* vitro functional analysis was
performed using the 293/EBNA1 cell line as described (Mol. Cell. Biol., 14: 5487-5494,
1994) and the pcDNA3Myc vector containing full-length *DNMT3B* coding sequences, and
p220.2 (Mol. Cell. Biol., 7: 379-387, 1987) as the assay plasmid. Cotransfected target DNA
was digested with the methylation-sensitive restriction endonuclease *HpaII*, and Southern
blot was performed using p220.2 as a probe. All of the transfections were done in duplicate
or triplicate for each experiment.

[0176] **Analysis of *DNTM3B* Sequence and *IGF2* Imprinting.** Direct PCR sequencing
of genomic DNA was performed to analyze the sequence of *DNMT3B*. All of the coding
exons published including exon-intron junctions were thoroughly examined. LOI of *IGF2*
was assessed according to hot-stop PCR (H. Uejima et al, Nat. Genet., 25: 375-376, 2000).

Results

[0177] **Hypomethylation of *H19* and *IGF2* DMRs in DNA Methyltransferase
Knockout Cell Lines.** HCT116 cells show normal imprinting but undergo LOI of *IGF2*
after somatic cell knockout of both *DNMT1* and *DNMT3B* (I. Rhee et al, Nature (Lond.),
416: 552-556, 2002). To test whether loss of methylation rather than the gain of
methylation is responsible for LOI in these cells, we examined directly the methylation of
two DMRs that distinguish parental alleles in human cells: the *H19* DMR 5-kb upstream of
H19 and methylated on the paternal allele (*H19* active, *IGF2* silent); and the *IGF2* DMR
within intron 2 of *IGF2* and methylated on the maternal allele (M.J. Sullivan et al, Ref.
Oncogene, 18: 7527-7534, 1999). Note that the DMRs in humans differ from the mouse, in
which there are three rather than one DMR within *IGF2* (H. Nakagawa et al, Proc. Natl.
Acad. Sci. USA, 98: 591-596, 2001). Bisulfite sequencing analysis of HCT116 cells, and
HCT116 cells lacking *DNMT1*, *DNMT3B*, or both, revealed that in the double-knockout
cells, which showed LOI, both the *H19* and *IGF2* DMRs were extensively hypomethylated
(Fig. 4). This hypomethylation was found in three separate double-knockout lines with LOI

and in none of single-knockout or wild-type lines with normal imprinting (Table 3) .

Table 3. IGF2 imprinting status and methylation alterations in DNMT knockout cell lines

		Methylation Status	
Cell Lines	**IGF2 LOI**	**H19 CBS1**	**IGF2 DMR**
WT[1]	No	Hyper	Half
T1KO-1	No	Hyper	Half
T1KO-2	No	Hyper	Half
3BKO-1	No	Hyper	Half
3BKO-2	No	Hyper	Half
DKO-1	Yes	Hypo	Hypo
DKO-2	Yes	Hypo	Hypo
DKO-3	Yes	Hypo	Hypo

[0178] **Hypomethylation of *H19* and *IGF2* DMRs in** Primary **CRCs.** To determine whether hypomethylation was also linked to LOI in primary colon cancers, we then analyzed 20 CRCs informative for imprinting status of *IGF2* (heterozygous for a transcribed polymorphism) by reverse transcription-PCR, 12 with LOI and 8 with normal imprinting. All 8 of the CRCs with normal imprinting showed the normal half-methylation pattern at the *IGF2* DMR, and all 12 of the CRCs with LOI showed marked hypomethylation of the *IGF2* DMR (P = 0.000007; Figs. 5 and 6). In tumors with normal imprinting, the fraction of CpG sites that were methylated was 43.6 ± 10.9%, whereas in tumors with LOI the fraction of sites methylated was 10.9 ± 9.4% (P < 0.0001). In addition, for each DMR, 15–20 clones were independently sequenced from the PCR product of each bisulfite-treated sample, and each experiment was repeated at least once. We also observed

---

[1] WT, HCT116 wild-type; T1KO, DNMT1 knockout; 3BKO, DNMT3B knockout; DKO, double-knockout; Half, normal half-methylation; Hypo, hypomethylation; Hyper, hypermethylation.

hypomethylation of the *H19* DMR in CRC, although the differences were not absolute as in the case of the *IGF2* DMR, but were in marked contrast to Wilms' tumors with LOI (Table 4). These results also differ markedly from those of Nakagawa et al. (H. Nakagawa et al, Proc. Natl. Acad. Sci. USA, 98: 591-596, 2001), who reported hypermethylation of CBS6 in colorectal cancer. Finally, because LOI is found at increased frequency in both tumor and normal tissue of patients with CRC, we also examined the matched normal mucosa of 3 CRC patients whose tumors showed LOI. As we reported earlier (H. Cui et al, Nat. Med., 4: 1276-280, 1998), the matched normal mucosa also showed LOI of *IGF2*, although methylation had not been examined in that study. We found the same pattern of hypomethylation in the normal colonic mucosa in each patient as we found in tumors (Table 4), indicating that this epigenetic abnormality was not limited to the cancers.

Table 4. *IGF2* imprinting status and methylation alterations in primary CRC.[2]

Sample No.	IGF2 LOI	Methylation Status		
		H19 CBS1	H19 CBS6	IGF DMR
1T	No	Half	Hypo	Half
2T	No	Half	Hypo	Half
3T	No	Half	Hypo	Half
4T	No	Half	Half	Half
5T	No	Half	Hypo	Half
6T	No	Half	Half	Half
7T	No	Half	Half	Half
8T	No	Hyper	Half	Half
9T	Yes	Hypo	Hypo	Hypo
10T	Yes	Hypo	Hypo	Hypo
11T	Yes	Hypo	Hypo	Hypo
12T	Yes	Hypo	Hypo	Hypo

[2] Annotation as in Table 1.

12N	Yes	Half	Hypo	Hypo
13T	Yes	Half	Hypo	Hypo
14T	Yes	Hypo	Hypo	Hypo
15T	Yes	Half	Hypo	Hypo
16T	Yes	Half	Half	Hypo
17T	Yes	Half	Hypo	Hypo
17N	Yes	Half	Hypo	Hypo
18T	Yes	Hypo	Hypo	Hypo
18N	Yes	Hypo	Hypo	Hypo
19T	Yes	Half	Half	Hypo
19N	Yes	Half	Half	Hypo
Fetus	No	Half	Half	Half

[0179]   **Neutral Polymorphisms of *DNMT3B* in Human CRCs.** Because LOI and hypomethylation were present in normal tissue, and *DNMT3B* appeared to play a role in LOI in HCT116 cells, we examined all 20 of the CRC for germ-line mutations in the *DNMT3B* gene. Six of 20 patients showed a single variation in the coding sequence leading to amino acid substitutions: G892T (G210W), G1390A (A376T), A1451G (Y396C), G2044A (V594I), G2086A (V608M), and T1436C (L391P). To distinguish between neutral and functional variants, we performed site-directed mutagenesis and transfection into 293/EBNA1 cells, together with an episomal vector, which was the target for *de novo* methylation. None of the variants disrupted *DNMT3B* methyltransferase activity (data not shown). Thus, these sequence variations represent neutral polymorphisms.

Discussion

[0180]   The study reported in this Example has two major results. First, hypomethylation, rather than hypermethylation, is linked to LOI of *IGF2* in human CRC based on two lines of evidence. In CRC lines in which hypomethylation is induced artificially by *DNMT1/DNMT3B* double knockout, LOI is found only in the hypomethylated lines. Indeed, unmodified HCT116 cells with hypermethylation of the *H19* DMR exhibit normal imprinting, even though Wilms' tumors with hypermethylation of the same sites

show LOI (27). Furthermore, in primary human CRC, as well, LOI is linked to hypomethylation rather than hypermethylation. The latter result is in contrast to the findings of Nakagawa *et al.* (35), who reported hypermethylation of the *H19* in CRC with LOI of *IGF2*. It should be remembered that the first epigenetic alterations found in human cancer was hypomethylation of DNA (1) and that CRC show global hypomethylation even in the presence of specific sites of increased DNA methylation (2). Furthermore, the assumption that CpG islands are universally hypomethylated is incorrect, as imprinted genes show normal methylation, and we have also identified recently many normally methylated CpG islands in normal cells (36). Therefore, a more correct and inclusive view is that cancers show epigenetic instability, including global hypomethylation, and sites of both aberrantly increased and decreased methylation, that lead to altered gene regulation.

[0181] The second major result of this study provided in this Example is that normal imprinting in the colon and LOI in CRC is specifically linked to the methylation status of a DMR within *IGF2* and not to a lesser extent to the methylation status of *H19*. Thus, all 8 of the cancers with normal imprinting showed normal half-methylation of the *IGF2* DMR and all 11 of the cancers showed hypomethylation of this DMR, as well as 3 matched normal mucosal specimens that also showed LOI. Takai *et al.* (37) recently described partial or complete hypomethylation of the *H19* ICR in two of four bladder cancers, but no relationship to *H19* imprinting; *IGF2* was not examined in that study. No alteration of *H19* imprinting was observed in the CRC examined here. It has been reported earlier that cancers with LOI also show LOI in the matched normal mucosa (Cui, H., Horon, I.L., Ohlsson, R., Hamilton, S.R., Feinberg, A.P. Loss of imprinting in normal tissue of colorectal cancer patients with microsatellite instability. Nat. Med., 4: 1276-280, 1998), so we would expect that this methylation abnormality is generally present in the colon of these cancer patients, as disclosed in Example 1.

[0182] An important implication of this result is that it suggests a mechanism for regulation of *IGF2* imprinting independent of enhancer competition. By the enhancer competition model, *IGF2* and *H19* promoters compete on the same chromosome for a shared enhancer, and access of the maternal *IGF2* allele to this enhancer is blocked by the

59

H19 DMR when unmethylated, likely because of the insulator activity of CTCF binding to the unmethylated *H19* DMR (22, 23, 24, 25, 26). However, in CRC with LOI, the *H19* DMR is hypomethylated on both alleles, and hypomethylation of the *IGF2* DMR is specifically linked to LOI of *IGF2* in both primary CRC and in HCT116 cells in which methyltransferases have been disrupted experimentally.

[0183]   Some clues to function are available from mouse studies, although it is difficult to relate mouse experiments precisely to the human, as the DMR sequences themselves differ between species. Nevertheless, the region corresponding to the human DMR studied here is in the same physical relationship to human *IGF2* exons 2 and 3, as is mouse "DMR0" to mouse *IGF2* pseudoexons 1 and 2 (34). To date, no mouse knockout of DMR0 by itself has been reported, although deletion of DMR1, or of DMR0 and DMR1 together, lead to activation of the normally silent maternal allele of *IGF2* (38, 39). The mouse knockout experiments suggest the existence of a transcriptional repressor within *IGF2* (38, 39). This hypothesis appears consistent with the results herein, and additionally the present results suggest that methylation of this human *IGF2* DMR recruits transcriptional repressors to the maternal allele. By this model, hypomethylation would lead to LOI by loss of association of these repressors to the *IGF2* DMR. The results herein also suggest two potentially valuable lines of experimentation: knockout of DMR0 in mouse and biochemical studies aimed at identifying factors of which the binding to the human *IGF2* DMR is lost in tumors with LOI.